

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: February 18, 2005, 15:30:38 ; Search time 139 Seconds

(without alignments)
795.737 Million cell updates/sec

Title: US-10-791-017A-2_COPY_319_656

Perfect score: 1922

Sequence: 1 GGRGGMGSGRGERGFRKPGG.....GGRGKDKGRHRRDRPY 338

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1380268 seqs, 327241040 residues

Total number of hits satisfying chosen parameters: 1380268

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications AA:*
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/2/pubpaa/US10D_PUBCOMB.pep:*
17: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
18: /cgn2_6/ptodata/2/pubpaa/US11_NEW_PUB.pep:*
19: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*
20: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1922	100.0	600	US-10-094-749-2713	Sequence 2713, App1
2	1922	100.0	656	US-10-791-017A-2	Sequence 2, App1
3	878	45.7	525	US-10-755-889-650	Sequence 650, App
4	799.5	41.6	260	US-10-408-765A-116	Sequence 116, App
5	623	32.4	156	US-09-925-301-1511	Sequence 1511, App
6	475	24.7	83	US-09-864-761-42397	Sequence 42397, A
7	423	22.0	467	US-10-437-963-121926	Sequence 121926, A
8	420	21.9	395	US-10-425-114-69744	Sequence 69744, A
9	319.5	16.6	578	US-10-437-963-164510	Sequence 164510, A
10	303	15.8	197	US-10-424-599-187688	Sequence 187688, A
11	303	15.8	1078	US-10-058-124-21	Sequence 21, App1
12	303	15.8	1466	US-10-402-089-4	Sequence 4, App1
13	303	15.8	1466	US-10-402-089-6	Sequence 6, App1

14	303	15.8	1466	US-10-402-072A-4	Sequence 4, App1
15	303	15.8	1466	US-10-402-072A-6	Sequence 6, App1
16	301	15.7	1466	US-09-918-715-226	Sequence 226, App
17	301	15.7	1466	US-10-177-293-68	Sequence 68, App1
18	301	15.7	1466	US-10-301-822-33	Sequence 33, App1
19	301	15.7	1466	US-10-257-021-72	Sequence 72, App1
20	301	15.7	1466	US-10-357-851-3	Sequence 3, App1
21	301	15.7	1466	US-10-358-024-3	Sequence 3, App1
22	301	15.7	1466	US-10-734-564-103	Sequence 103, App
23	300.5	15.6	1466	US-10-402-089-12	Sequence 12, App1
24	300.5	15.6	1466	US-10-402-072A-12	Sequence 120427
25	284.5	14.8	462	US-10-437-963-120427	Sequence 70, App1
26	275.5	14.3	1496	US-10-177-293-70	Sequence 35, App1
27	275.5	14.3	1496	US-10-301-822-35	Sequence 74, App1
28	275.5	14.3	1496	US-10-236-031B-74	Sequence 22, App1
29	275.5	14.3	1496	US-10-468-091-22	Sequence 248, App
30	275.5	14.3	1496	US-10-788-792-248	Sequence 45, App1
31	272.5	14.2	651	US-10-488-056-45	Sequence 8, App1
32	272.5	14.2	1449	US-10-402-089-8	Sequence 8, App1
33	272.5	14.2	1449	US-10-402-072A-8	Sequence 8, App1
34	271.5	14.1	968	US-10-291-172-739	Sequence 739, App
35	271.5	14.1	968	US-10-221-278-739	Sequence 739, App
36	270.5	14.1	1046	US-10-156-761-10088	Sequence 10088, A
37	270	14.0	1487	US-10-468-091-6	Sequence 6, App1
38	270	14.0	1497	US-10-468-091-23	Sequence 23, App1
39	268.5	14.0	386	US-10-408-765A-959	Sequence 959, App
40	268.5	14.0	1463	US-10-402-089-2	Sequence 2, App1
41	268.5	14.0	1463	US-10-402-072A-2	Sequence 2, App1
42	267	13.9	396	US-10-424-599-251915	Sequence 251915
43	266.5	13.9	940	US-10-291-172-363	Sequence 363, App
44	266.5	13.9	940	US-10-221-278-363	Sequence 363, App
45	265.5	13.8	1453	US-10-468-091-26	Sequence 26, App1

ALIGNMENTS

RESULT 1
US-10-094-749-2713
; Sequence 2713, Application US/10094749
; Publication No. US20030219741A1
GENERAL INFORMATION:
; APPLICANT: ISOGAI, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
; APPLICANT: WAKAMATSU, AI
; APPLICANT: SATO, HIROYUKI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: YAMAMOTO, JUN-ICHI
; APPLICANT: ISONO, YUUKO
; APPLICANT: HIO, YURI
; APPLICANT: OTSUKA, KAORU
; APPLICANT: NAGAI, KEIICHI
; APPLICANT: NAGAI, KEIICHI
; APPLICANT: TAMECHINA, ICHIRO
; APPLICANT: SEKI, NAOHITO
; APPLICANT: YOSHIKAWA, TSUTOMU
; APPLICANT: OTSUKA, MOTOTYUKI
; APPLICANT: NAGAHARA, KENJI
; APPLICANT: MASUHO, YASUHIKO
TITLE OF INVENTION: NOVEL FULL-LENGTH CDNA
; FILE REFERENCE: 084335/0160
; CURRENT APPLICATION NUMBER: US/10/094,749
; PRIOR FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: 60/350,435
; PRIOR FILING DATE: 2002-01-24
; PRIOR APPLICATION NUMBER: JP 2001-328381
; NUMBER OF SEQ ID NOS: 3381
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 2713
; LENGTH: 600
; TYPE: PRT

ORGANISM: Homo sapiens
US-10-094-749-2713

Query Match 100.0%; Score 1922; DB 15; Length 600;
Best Local Similarity 100.0%; Pred. No. 3.1e-128;
Matches 338; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 GGRGMSAGERGKPKGPMDEBPDLDGPPVDPEDSDNSAIYVQGLNDSVTLLDILA 60
263 GGRGMSAGERGKPKGPMDEBPDLDGPPVDPEDSDNSAIYVQGLNDSVTLLDILA 322
61 DFFKQGVVNMNRKTQPMIHYLDKETGPKGDATVSYEDPTAKAAVEMFGKDFQGS 120
323 DFFKQGVVNMNRKTQPMIHYLDKETGPKGDATVSYEDPTAKAAVEMFGKDFQGS 382
121 KLVSLARKKPPMNSMRGGLPREGRMPPPLRGPGGPGGPGMGRGGRGDRGFP 180
383 KLVSLARKKPPMNSMRGGLPREGRMPPPLRGPGGPGGPGMGRGGRGDRGFP 442
181 PRGPRGSRGNPSGGVVOHRAQDMOCNPGCCNONPAMRTECNQCAKPKREGFLPPFP 240
443 PRGPRGSRGNPSGGVVOHRAQDMOCNPGCCNONPAMRTECNQCAKPKREGFLPPFP 502
241 PGDGRGGRGPGKRGGRGLMDRGPGGMFRGGRGGRGGRGGRGGRGGRGGRG 300
503 PGDGRGGRGPGKRGGRGLMDRGPGGMFRGGRGGRGGRGGRGGRGGRGGRG 562
301 GGPPLMEQMGRRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGR 338
563 GGPPLMEQMGRRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGR 600

RESULT 2

US-10-791-017a-2
Sequence 2, Application US/10791017A
Publication No. US20040197827A1
GENERAL INFORMATION:
APPLICANT: JENAPHARM GmbH & Co. KG
TITLE OF INVENTION: Methods for Determining Hormonal Effects of Substances
FILE REFERENCE: Pat 3684/11
CURRENT APPLICATION NUMBER: US/10/791,017A
NUMBER OF SEQ ID NOS: 7
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 2
LENGTH: 656
TYPE: PRT
ORGANISM: Homo sapiens
US-10-791-017a-2

Query Match 100.0%; Score 1922; DB 16; Length 656;
Best Local Similarity 100.0%; Pred. No. 3.3e-128;
Matches 338; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 GGRGMSAGERGKPKGPMDEBPDLDGPPVDPEDSDNSAIYVQGLNDSVTLLDILA 60
319 GGRGMSAGERGKPKGPMDEBPDLDGPPVDPEDSDNSAIYVQGLNDSVTLLDILA 378
61 DFFKQGVVNMNRKTQPMIHYLDKETGPKGDATVSYEDPTAKAAVEMFGKDFQGS 120
379 DFFKQGVVNMNRKTQPMIHYLDKETGPKGDATVSYEDPTAKAAVEMFGKDFQGS 438
121 KLVSLARKKPPMNSMRGGLPREGRMPPPLRGPGGPGGPGMGRGGRGDRGFP 180
439 KLVSLARKKPPMNSMRGGLPREGRMPPPLRGPGGPGGPGMGRGGRGDRGFP 498
181 PRGPRGSRGNPSGGVVOHRAQDMOCNPGCCNONPAMRTECNQCAKPKREGFLPPFP 240
499 PRGPRGSRGNPSGGVVOHRAQDMOCNPGCCNONPAMRTECNQCAKPKREGFLPPFP 558
241 PGDGRGGRGPGKRGGRGLMDRGPGGMFRGGRGGRGGRGGRGGRGGRGGRG 300
559 PGDGRGGRGPGKRGGRGLMDRGPGGMFRGGRGGRGGRGGRGGRGGRGGRG 618

QY 301 GGPPLMEQMGRRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGR 338
DB 619 GGPPLMEQMGRRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGR 656

RESULT 3

US-10-755-889-650
Sequence 650, Application US/10755889
Publication No. US20040171823A1
GENERAL INFORMATION:
APPLICANT: Bristol-Myers Squibb Company
TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES ASSOCIATED WITH THE NF-KB
FILE REFERENCE: D0284 NP
CURRENT APPLICATION NUMBER: US/10/755,889
CURRENT FILING DATE: 2004-01-13
PRIOR APPLICATION NUMBER: U.S. 60/440,068
PRIOR FILING DATE: 2003-01-14
PRIOR APPLICATION NUMBER: U.S. 60/469,757
PRIOR FILING DATE: 2003-05-12
SOFTWARE: PatentIn version 3.2
SEQ ID NO 650
LENGTH: 525
TYPE: PRT
ORGANISM: Homo sapiens
US-10-755-889-650

Query Match 45.7%; Score 878; DB 16; Length 525;
Best Local Similarity 52.9%; Pred. No. 2.9e-54;
Matches 182; Conservative 36; Mismatches 54; Indels 72; Gaps 13;

1 GGRGMSAGERGKPKGPMDEBPDLDGPPVDPEDSDNSAIYVQGLNDSVTLLDILA 60
248 GGRGMSAGERGKPKGPMDEBPDLDGPPVDPEDSDNSAIYVQGLNDSVTLLDILA 301
61 DFFKQGVVNMNRKTQPMIHYLDKETGPKGDATVSYEDPTAKAAVEMFGKDFQGS 120
302 DFFKQGVVNMNRKTQPMIHYLDKETGPKGDATVSYEDPTAKAAVEMFGKDFQGS 361
121 KLVSLARKKPPMNSMRGGLPREGRMPPPLRGPGGPGGPGMGRGGRGDRGFP 180
362 KLVSLARKKPPMNSMRGGLPREGRMPPPLRGPGGPGGPGMGRGGRGDRGFP 410
181 PRGPRGSRGNPSGGVVOHRAQDMOCNPGCCNONPAMRTECNQCAKPKREGFLPPFP 240
411 PRGPRGSRGNPSGGVVOHRAQDMOCNPGCCNONPAMRTECNQCAKPKREGFLPPFP 453
241 PGDGRGGRGPGKRGGRGLMDRGPGGMFRGGRGGRGGRGGRGGRGGRGGRG 295
454 PGDGRGGRGPGKRGGRGLMDRGPGGMFRGGRGGRGGRGGRGGRGGRGGRG 505
296 RRGPGGPGPLMEQMGRRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRG 338
506 RRGPGGPGPLMEQMGRRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRG 525

RESULT 4

US-10-408-765A-116
Sequence 116, Application US/10408765A
Publication No. US20040101874A1
GENERAL INFORMATION:
APPLICANT: Ghosh, Soumitra S.
APPLICANT: Fahy, Eoin D.
APPLICANT: Zhang, Bing
APPLICANT: Gibson, Bradford W.
APPLICANT: Taylor, Steven W.
APPLICANT: Glenn, Gary M.
APPLICANT: Warnock, Dale E.
TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
TITLE OF INVENTION: IDENTIFIED IN THE MITOCHONDRIAL PROTEOME
FILE REFERENCE: 660088.465

CURRENT APPLICATION NUMBER: US/10/408,765A
CURRENT FILING DATE: 2003-04-04
NUMBER OF SEQ ID NOS: 3077
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 116
LENGTH: 260
TYPE: PR1
ORGANISM: Homo sapiens
US-10-408-765A-116

Query Match 41.6%; Score 799.5; DB 16; Length 260;
Best Local Similarity 51.4%; Pred. No. 5,3e-49;
Matches 167; Conservative 34; Mismatches 53; Indels 71; Gaps 12;

QY 20 GPMDEBDDLDGPPVDEDDSDNSAIYVGLINDSVTLDDIADPFKQCVVKNKRRTGQPM 79
DB 1 GPRDQGSRRH-----SEQDNSDNTTIFVQGLGEVITLESVADYFKQIGIKTKWKTGQPM 55
QY 80 IHLYLDKGTGKPRGDATVSEDPPTAKAWEVPGDQFQSGKLVSLARKKPPMNSMRGG 139
DB 56 INLYTDETKLKGSEATVSPDDPSAVALDWPDGKFSGNPLKVSFATRRADFN--RGG 113
QY 140 LPPREGGMPPLRGCGPGCGPGGPMGRGGRGDRGPPRGSRGNGNPGGQNVQH 199
DB 114 GNGRGGRG-----RGGMGBGGYGGG-----GSGGGGGRGFPSSG-----GGGGGQ 155
QY 200 RAGDMCCNPGCCANONFAMTECNQCAPPEEGFLPPPPPGGDRGGRGPGGMR--GGRG 258
DB 156 RAGDMCCNPGTCNNMFSWNECNQCAPKFDG-----PG-----GGPGGSHMGANY 202
QY 259 GLMDRGPGGMPRG-----GSGGDRGGRGGR--GMDRGFGGGRGGRGPPGPMLEQMG 314
DB 203 GDRRGGRGGYDRGTRGRGDRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGR 240
QY 315 RGRGGRGGRGMD--KGEHRQERRDRPY 338
DB 241 -----GPGKMDSRGHRQDRRERY 260

RESULT 5
US-09-925-301-1511
Sequence 1511, Application US/09925301
Patent No. US20020052308A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
FILE REFERENCE: PA106
CURRENT APPLICATION NUMBER: US/09/925,301
PRIOR APPLICATION NUMBER: PCT/US00/05682
PRIOR FILING DATE: 2001-08-10
PRIOR FILING DATE: 2000-03-08
PRIOR APPLICATION NUMBER: 60/124,270
PRIOR FILING DATE: 1999-03-12
NUMBER OF SEQ ID NOS: 1694
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1511
LENGTH: 156
TYPE: PR1
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE
LOCATION: (11)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
LOCATION: (104)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (143)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-301-1511

Query Match 32.4%; Score 623; DB 9; Length 156;
Best Local Similarity 73.2%; Pred. No. 1e-36;

Matches 115; Conservative 1; Mismatches 7; Indels 34; Gaps 4;
QY 160 GGFPGPMGRMGSGDRGCGFPFPPRGSRGNGPSSGQNVQHRADWCCPNP----- 209
DB 18 GSPG--LQERGTR--DRGFPFPPRGSRGNGPSSGQNVQHRADWCCPNPSIDFCCDVI 73
QY 210 ---GCCNONFAMTECNQCAPPEEGFLPPPPPGGDRGGRGGRGGRGGRGGRGGR 266
DB 74 VCRCCANONFAMTECNQ-----GDRGGRGGRGGRGGRGGRGGRGGRGGR 116
QY 267 GGMFRGGRGDRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGR 303
DB 117 GGMFRGGRGDRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGR 153

RESULT 6
US-09-864-761-42397
Sequence 42397, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aeomica-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
PRIOR FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 42397
LENGTH: 83
TYPE: PR1
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC000026.3

OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.2
 OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.9
 OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.3
 OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2
 OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.5
 OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.5
 OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 5.8
 OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.7
 US-09-864-761-42397

Query Match 24.7%; Score 475; DB 9; Length 83;
 Best Local Similarity 100.0%; Pred. No. 1.7e-26;
 Matches 83; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 243 GDRGRGPGMGRGGLMDRGGPGMFRGGRGDRGFRGGRGMDRGGRRGGPG 302
 DB 1 GDRGRGPGMGRGGLMDRGGPGMFRGGRGDRGFRGGRGMDRGGRRGGPG 60
 QY 303 PPGPLMEQMGRRGRRGGPGKMD 325
 DB 61 PPGPLMEQMGRRGRRGGPGKMD 83

RESULT 7
 US-10-437-963-191926

Sequence 191926, Application US/10437963
 Publication No. US20040123343A1

GENERAL INFORMATION:
 APPLICANT: La Rosa, Thomas J.
 APPLICANT: Kovalic, David K.
 APPLICANT: Zhou, Yihua
 APPLICANT: Cao, Yongwei
 APPLICANT: Wu, Wei
 APPLICANT: Boukharov, Andrey A.
 APPLICANT: Barbazuk, Brad
 TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
 TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
 FILE REFERENCE: 38-21(53221)B
 CURRENT APPLICATION NUMBER: US/10/437,963
 CURRENT FILING DATE: 2003-05-14
 NUMBER OF SEQ ID NOS: 204966
 SEQ ID NO 191926
 LENGTH: 467
 TYPE: PRT
 ORGANISM: Oryza sativa
 FEATURE:
 OTHER INFORMATION: Clone ID: PAT_MRT4530_881C.1.pep
 US-10-437-963-191926

Query Match 22.0%; Score 423; DB 16; Length 467;
 Best Local Similarity 32.4%; Pred. No. 4.9e-22;
 Matches 113; Conservative 29; Mismatches 77; Indels 130; Gaps 12;

QY 31 GPPVDPEDSDNSAIYQGLNDSVTLDDLADFFKQCGVYKANKRTGPMIHYLDKETGK 90
 DB 95 GPP-----NSVYVCNLPFGTDEMLADYFETGLKKDKRTGPKMITYRDKYTNE 146
 QY 91 PKGDATVSYEDPPTAKAAYEMFDGKDFGSKLKVSLARKK-----PPMNSMRGGLPP 142
 DB 147 PKGDATVSYEDPPTAKAAYEMFDGKDFGSKLKVSLARKK-----PPMNSMRGGLPP 202
 QY 143 REGRMPPPLRGGRGCGGPGMGRGGR-----GGDRGCFPPRGGRGSRGNPSGGGV 197
 DB 203 -----LGQDELNDNGAGRG-----RGHGDPGKA 226
 QY 198 QHRAGDMQCPNPGCGNQNFAMRTECNQCKAPKPEGFLPPFPFGGDRGGRGGRGGR 257
 DB 227 WQDGDMLCINTSCGVNFAFRGVCNRCGAARPG----- 261
 QY 258 GGLMDRGGPGMGRGGRGDRGFRGGRGMDRGGFGGRRGGRGPGPG-----PLM 308
 DB 262 ---VSGSGAGG---GGRG-----RGRGSD-DAKGSRAAAYGPGFLGPPMDWGCPMC 307

QY 309 EQM-----GRRGRGGRGPGK-MDKGRORRRDR 336
 DB 308 GNINMAKMKNCINTTKRGNHEGVRGGGGYKELDEBELEEVYKRR 356

RESULT 8
 US-10-425-114-69744
 Sequence 69744, Application US/10425114
 Publication No. US20040034888A1

GENERAL INFORMATION:
 APPLICANT: Liu, Jindong
 APPLICANT: Zhou, Yihua
 APPLICANT: Kovalic, David K.
 APPLICANT: Screen, Steven E
 APPLICANT: Tabaska, Jack E
 APPLICANT: Cao, Yongwei
 TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
 TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
 FILE REFERENCE: 38-21(53313)B
 CURRENT APPLICATION NUMBER: US/10/425,114
 CURRENT FILING DATE: 2003-04-28
 NUMBER OF SEQ ID NOS: 73128
 SEQ ID NO 69744
 LENGTH: 395
 TYPE: PRT
 ORGANISM: Zea mays
 FEATURE:
 OTHER INFORMATION: Clone ID: UC-ZMFLB73242A07_FLI.pep
 US-10-425-114-69744

Query Match 21.9%; Score 420; DB 15; Length 395;
 Best Local Similarity 32.4%; Pred. No. 6.7e-22;
 Matches 113; Conservative 31; Mismatches 95; Indels 110; Gaps 11;

QY 20 GPMDEGPDLDLGPVDPEDSDNSAIYQGLNDSVTLDDLADFFKQCGVYKANKRTGPM 79
 DB 12 GPMDEGPDLDLGPVDPEDSDNSAIYQGLNDSVTLDDLADFFKQCGVYKANKRTGPM 57
 QY 80 IHIYDKETGKRGDATTYEDPPTAKAAYEMFDGDFGSKLKVSLARKKPP----- 132
 DB 58 IHIYDKETGKRGDATTYEDPPTAKAAYEMFDGDFGSKLKVSLARKKPP----- 117
 QY 133 -MNSMRGGLPPREGRMPPPLRGGRGCGGPGMGRGGRGDRGFRGGRGSRGNP 191
 DB 118 -MNSMRGGLPPREGRMPPPLRGGRGCGGPGMGRGGRGDRGFRGGRGSRGNP 142
 QY 192 SGGGVQHRAGDMQCPNPGCGNQNFAMRTECNQCKAPKPEGFLPPFPFGGDRGGRGPG 251
 DB 143 SGGGVQHRAGDMQCPNPGCGNQNFAMRTECNQCKAPKPEGFLPPFPFGGDRGGRGPG 188
 QY 252 GMRGGRGGLMDRGGPGMGRG--GRGDRGFRGGRGMDRGGRRG--GPGGPGPLM 308
 DB 189 GMRGGRGGLMDRGGPGMGRG--GRGDRGFRGGRGMDRGGRRG--GPGGPGPLM 233
 QY 309 EQM-----GRRGRGGRGPGK-MDKGRORRRDR 336
 DB 234 GNINMAKMKNCINTSKRPTNEGVRGGGRGGYKELDEBELEEVYKRR 262

RESULT 9
 US-10-437-963-164510
 Sequence 164510, Application US/10437963
 Publication No. US20040123343A1
 GENERAL INFORMATION:
 APPLICANT: La Rosa, Thomas J.
 APPLICANT: Kovalic, David K.
 APPLICANT: Zhou, Yihua
 APPLICANT: Cao, Yongwei
 APPLICANT: Wu, Wei
 APPLICANT: Boukharov, Andrey A.
 APPLICANT: Barbazuk, Brad
 APPLICANT: Li, Ping


```

/ TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated with
/ FILE REFERENCE: 38-21(5322)B
/ CURRENT APPLICATION NUMBER: US/10/437,963
/ NUMBER OF SEQ ID NOS: 204966
/ SEQ ID NO 164510
/ LENGTH: 578
/ TYPE: PRT
/ ORGANISM: Oryza sativa
/ FEATURE:
/ OTHER INFORMATION: Clone ID: PAT_MRT4530_63401C.1.pep
US-10-437-963-164510

Query Match          16.6%; Score 319.5; DB 16; Length 578;
Best Local Similarity 3.5%; Pred. No. 1.3e-14;
Matches 90; Conservative 9; Mismatches 74; Indels 67; Gaps 8;

QY 145 GRGMPPLRGGPGGPGGPGMGRGGRG-----DRGPPRRGPRGRGNPSGGGNVQ 198
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 64 GCGVGGCGGGCGGGCGGGGGGGYGG-GRGGGGGGGGYGGGGGGGGGGGGGGGGGG 122

QY 199 HRAADMCCNP-----GCCGNQ 215
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 123 GRDGDWVCPDPRILHGACVCKPTKFLFHRGINIRLQFHFSSNIPNFVYHSSCGNVN 182

QY 216 FAARTENQCKAKPEGFLEPPPPPGCGDGRGPGGCGGLMDRGGPGGKFRGGRG 275
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 183 FAARTENQCKAKPSAPG-----GGGGGGGGGGYKSGGGGGYVRGG-GRFSSGGGGG 232

QY 276 G-DRGG---FRGGRGMDRGGFGGGRRGGPGGPPGLMEQNGRRGGRGGGGKXDKGHRQ 331
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 233 GYVRGGGDYVSGRG---GRTGGGGGGGGYVRGGGDDRGFDHRGGGGYGGRDQGNQR 289

```

```

RESULT 10
US-10-424-599-187688
/ Sequence 187688, Application US/10424599
/ Publication NO. US20040031072A1
/ GENERAL INFORMATION:
/ APPLICANT: La Rosa Thomas J
/ APPLICANT: Kovalic David K
/ APPLICANT: Zhou Yihua
/ APPLICANT: Cao Yongwei
/ TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with
/ FILE REFERENCE: 38-21(53223)B
/ CURRENT APPLICATION NUMBER: US/10/424,599
/ CURRENT FILING DATE: 2003-04-28
/ NUMBER OF SEQ ID NOS: 285684
/ SEQ ID NO 187688
/ LENGTH: 197
/ TYPE: PRT
/ ORGANISM: Glycine max
/ FEATURE:
/ NAME/KEY: unsure
/ LOCATION: (1)..(197)
/ OTHER INFORMATION: unsure at all Xaa locations
/ FEATURE:
/ OTHER INFORMATION: Clone ID: PAT_MRT3847_140496C.1.pep
US-10-424-599-187688

```

```

Query Match          15.8%; Score 303; DB 15; Length 197;
Best Local Similarity 34.3%; Pred. No. 6.6e-14;
Matches 79; Conservative 21; Mismatches 70; Indels 60; Gaps 7;

QY 42 NSAIYQGLSDSTLDD--LADFFKQCGVYKMKRTGQPMIHIYLDKRTGKPKGDATVSY 99
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 12 NSGVY--CNLPYGTDDIMLAIEFTIGLTKDKRTGRPTIMLRKKTETNPPKGDATVY 69

QY 100 EDPPTAKAAVEMEDGDFQSGKLKVLARKK-----PPMSMRGGLPPREGRG 147
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 70 EDPHAAVAALVEMENNDFTGNTIGTIVFIASKKNDBQAYNAAVEPVVADVGL-----EE 124

```

```

QY 148 MPEPLRGPGGPGGPGMGRGGRGGDRGGFPFPRGRGRGNPSGGGNVQHRAGDWCP 207
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 125 TTKDVGGSG-----RGXXQND-SGXWQDGDWICL 156

QY 208 NPGCGNONFAMRTENQCKAKPEGFLEPPPPPGCGDRGRGGPGKMRGGR 257
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 157 NTGCSNVNFAFRGACNRXGTAR-----SAGASGISGAGRGKR 195

```

```

RESULT 11
US-10-058-124-21
/ Sequence 21, Application US/10058124
/ Publication NO. US20030119058A1
/ GENERAL INFORMATION:
/ APPLICANT: Qvist, Per
/ APPLICANT: Bonde, Martin
/ TITLE OF INVENTION: A Method for Assaying Collagen Fragments
/ in Body Fluids, A Test Kit and Means for Carrying Out the
/ Method and Use of the Method to Diagnose the Presence of
/ Disorders Associated with the Metabolism of

NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSER: Darby & Darby PC
STREET: 805 Third Avenue
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10022

COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/058,124
FILING DATE: 29-Jan-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/570,573
FILING DATE: 2002-MAY-12
APPLICATION NUMBER: 08/187,319
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Gogoris, Adda C
REGISTRATION NUMBER: 29,714
REFERENCE/DOCKET NUMBER: 4305/08701
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-527-7700
TELEFAX: 212-753-6237
TELEX: 236687

INFORMATION FOR SEQ ID NO: 21:
SEQUENCE CHARACTERISTICS:
LENGTH: 1078 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
IMMEDIATE SOURCE:
CLONE: COLLAGEN ALPHA 1 (III)
SEQUENCE DESCRIPTION: SEQ ID NO: 21:
US-10-058-124-21

```

```

Query Match          15.8%; Score 303; DB 14; Length 1078;
Best Local Similarity 30.3%; Pred. No. 3.7e-13;
Matches 115; Conservative 15; Mismatches 126; Indels 124; Gaps 16;

QY 1 GRRGGSAGSRGGRGKPGPMDEGPDLDGPPVDPEEDSDNSAIYVQGLSDSTLDDLA 60
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 291 GAKGSGRGRGERGEGIPGVPAKGEDKDGSPGPGANGLGGAAGERALDS----- 343

QY 61 DFFKQCGVYKMKRTGQPMIHIYLDKRTGKPKGDATVSYEDPPTAKAAVEMEDGDFQGS 120

```

Db 344 -----RGPAGPNCIPGEKGPAGS----- 361
 QY 121 KUKVSLARKKPPNNNSMRG--GLPPEGRGMPPLRGCGPGGPG--GPMGRMG--G 171
 Db 362 -----RGAPGACGAGAAAGEPGRDGPVCGPGRGMPGSPGCGPBGDGRPGPSGSEGS 414
 QY 172 RCGDRGPPRRGRGRSG-----NPSGGGVHQRADDMQCPNPGCNQNPAMTECNQCK 226
 Db 415 RPGRPG--PSGRGPGVWGFPGPKNDGAPGKNGERGGPG--GPGPGQ----- 459
 QY 227 APKEGEFLPPFPFPP-----PGDGRGRGPGGMRG-----GRGIMDR----- 263
 Db 460 PPKKNBEYRQGGPPTGPGGDKDTPRGPRQGLQGLPTGSGPGENEKGEPKGEANG 519
 QY 264 --GGPGMFRGGRGDRG-----GFRGGRMDRGFGGRRGPGGPG--PLME 309
 Db 520 APGAPGKGDAAGAPGERGPPGLAGAPGLRGAG--PGPEGKGAAGPPGPPGAATPGLQ 578
 QY 310 QMGRRGGRGPG--KMDKGE 328
 Db 579 GMPERGGLGSPGPKDKGE 598

RESULT 12

US-10-402-089-4
 ; Sequence 4, Application US/10402089
 ; Publication No. US20040005663A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Bell, Marcum P.
 ; APPLICANT: Neff, Thomas B.
 ; APPLICANT: Polarek, James W.
 ; APPLICANT: Sealey, Todd W.
 ; TITLE OF INVENTION: PORCINE COLLAGENS AND GELATINS
 ; FILE REFERENCE: PP0402.3 CON
 ; CURRENT APPLICATION NUMBER: US/10/402,089
 ; CURRENT FILING DATE: 2003-03-26
 ; PRIOR APPLICATION NUMBER: US 09/709,700
 ; PRIOR FILING DATE: 2000-11-10
 ; NUMBER OF SEQ ID NOS: 72
 ; SOFTWARE: PatentIn version 3.2
 ; SEQ ID NO 4
 ; LENGTH: 1466
 ; TYPE: PRT
 ; ORGANISM: Bos Taurus
 ; US-10-402-089-4

Query Match 15.8%; Score 303; DB 15; Length 1466;
 Best Local Similarity 31.2%; Pred. No. 5.1e-13;
 Matches 130; Conservative 22; Mismatches 157; Indels 108; Gaps 21;
 QY 1 GGRGMSAGERGGRGFMDEGPDLDLGPVPDPEDSDNSAIYVQGLNDSVTLDDL 60
 Db 348 GAKGEVGPAGSPGSSGAPGGRGEPGSHAGAPPPGPPSGNSPGKGMGPAGI PGAP 407
 QY 61 DFFKQGVVKNKRTGQPMIHYLDK-----ETGK--PKGDATVSYE-----DPTAKA 108
 Db 408 -----GLIGARGPPEPTNGVPGQKGAAGEPKNGAKADPPGRGERGAGSBGLAGPK 461
 QY 109 VEWFDGKD-----FOGSKLVSLARKKPPNNNSMRG----- 138
 Db 462 GE--DGKDSRPERGANGLPGAAGERVPPRGPAGANGLPGEKGP GDDGGGPAPARG 519
 QY 139 --GLPPEGRGMPPLRGSGPGGPG--GPMGRMG-----GRGDRG-----G 178
 Db 520 VAGPGRDGLPGGGLRGLIGSPGSPGSDKPKPPGSGQETGRPPGSPGPPGPGVWG 579
 QY 179 FPPRGRGSRGNPS-----GGGVNQHRAAGDMQCPNPGCGNPFAMRTECNQCKA--- 227
 Db 580 FP--GPKNDGAPGKNGERGGPGGPGG--KNGETGPPGPPGPTGSPGDKDGTG 633
 QY 228 -PKPEGF--LPPFPFPPG-----GDRGGRGPGMRGGRGIMDRGCGMFRGGRG 276

Db 634 PPGQGLQGLPTGSPGPGENGKGEPPKGEAGAPGIPGKGC---DSGAPGRGPPGAG 690
 QY 277 DRGFRGRGMDRGFRGGRGPGGPG--PLMEQVGRGRGGRGPG--KMDKGE 328
 Db 691 PPGP--RGAG--PPGEGGKGAAGPPGPPGSPGACTPGLQGMPPRRGGPGGPGKDKGE 745

RESULT 13

US-10-402-089-6
 ; Sequence 6, Application US/10402089
 ; Publication No. US20040005663A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Bell, Marcum P.
 ; APPLICANT: Neff, Thomas B.
 ; APPLICANT: Polarek, James W.
 ; APPLICANT: Sealey, Todd W.
 ; TITLE OF INVENTION: PORCINE COLLAGENS AND GELATINS
 ; FILE REFERENCE: PP0402.3 CON
 ; CURRENT APPLICATION NUMBER: US/10/402,089
 ; CURRENT FILING DATE: 2003-03-26
 ; PRIOR APPLICATION NUMBER: US 09/709,700
 ; PRIOR FILING DATE: 2000-11-10
 ; NUMBER OF SEQ ID NOS: 72
 ; SOFTWARE: PatentIn version 3.2
 ; SEQ ID NO 6
 ; LENGTH: 1466
 ; TYPE: PRT
 ; ORGANISM: Sus scrofa
 ; US-10-402-089-6

Query Match 15.8%; Score 303; DB 15; Length 1466;
 Best Local Similarity 31.2%; Pred. No. 5.1e-13;
 Matches 130; Conservative 22; Mismatches 157; Indels 108; Gaps 21;
 QY 1 GGRGMSAGERGGRGFMDEGPDLDLGPVPDPEDSDNSAIYVQGLNDSVTLDDL 60
 Db 348 GAKGEVGPAGSPGSSGAPGGRGEPGSHAGAPPPGPPSGNSPGKGMGPAGI PGAP 407
 QY 61 DFFKQGVVKNKRTGQPMIHYLDK-----ETGK--PKGDATVSYE-----DPTAKA 108
 Db 408 -----GLIGARGPPEPTNGVPGQKGAAGEPKNGAKADPPGRGERGAGSBGLAGPK 461
 QY 109 VEWFDGKD-----FOGSKLVSLARKKPPNNNSMRG----- 138
 Db 462 GE--DGKDSRPERGANGLPGAAGERVPPRGPAGANGLPGEKGP GDDGGGPAPARG 519
 QY 139 --GLPPEGRGMPPLRGSGPGGPG--GPMGRMG-----GRGDRG-----G 178
 Db 520 VAGPGRDGLPGGGLRGLIGSPGSPGSDKPKPPGSGQETGRPPGSPGPPGPGVWG 579
 QY 179 FPPRGRGSRGNPS-----GGGVNQHRAAGDMQCPNPGCGNPFAMRTECNQCKA--- 227
 Db 580 FP--GPKNDGAPGKNGERGGPGGPGG--KNGETGPPGPPGPTGSPGDKDGTG 633
 QY 277 DRGFRGRGMDRGFRGGRGPGGPG--PLMEQVGRGRGGRGPG--KMDKGE 328
 Db 691 PPGP--RGAG--PPGEGGKGAAGPPGPPGSPGACTPGLQGMPPRRGGPGGPGKDKGE 745

RESULT 14

US-10-402-072A-4
 ; Sequence 4, Application US/10402072A
 ; Publication No. US20040018592A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Bell, Marcum P.
 ; APPLICANT: Neff, Thomas B.
 ; APPLICANT: Polarek, James W.
 ; APPLICANT: Sealey, Todd W.
 ; TITLE OF INVENTION: BOVINE COLLAGENS AND GELATINS

FILE REFERENCE: PP0402.2 CON
 CURRENT APPLICATION NUMBER: US/10/402,072A
 CURRENT FILING DATE: 2003-03-26
 PRIOR APPLICATION NUMBER: US 09/709,700
 PRIOR FILING DATE: 2000-11-10
 NUMBER OF SEQ ID NOS: 72
 SOFTWARE: PatentIn version 3.2
 SEQ ID NO 4
 LENGTH: 1466
 TYPE: PRT
 ORGANISM: Bos Taurus
 US-10-402-072A-4

Query Match 15.8%; Score 303; DB 15; Length 1466;
 Best Local Similarity 31.2%; Pred. No. 5.1e-13;
 Matches 130; Conservative 22; Mismatches 157; Indels 108; Gaps 21;

QY 1 GGRGMSAGEGRGFKRGGPMDEGPDLDGPPVDPEDSDNSAIYVQGLNDSTVTLDDLA 60
 DB 348 GAKGEVGPASSPGSSGAPGQGRGEPFGHAGAPPPPPSGNSPGKGEKMPAGIPGAP 407
 QY 61 DFFKQCGVVMKRRKTGQPMIHLYDK-----ETGK--PKGDATVSYE-----DPTAKAA 108
 DB 408 -----GLIGARPPPPGPTGTVGQGRGAAGEPGKNAKDDPPRGRGERGASFGIAGPK 461
 QY 109 VEMFDGKD-----FQSKLVSLARKKPPMNSMRG----- 138
 DB 462 GE--DGKDGSPGEBGANGLPGAAGERGVPPFRGPAGANGLPGEKGPDRGPGPAPRG 519
 QY 139 --GLPPREGMPPPLRGSGPGPGPG-----GPMGRMG-----GRGDDG-----G 178
 DB 520 VAEPPGIDGLPGGDLGRLGIPSGPGSDGKPPGSGQGETGRPPGSPGPPGQPGVWG 579
 QY 179 PPRGPRGSRGNPS-----GGNVQHRAGDMQCPNPGCGNQNFAMRTECNQCKA--- 227
 DB 580 FP--GPKNDGAPKNGERGGPGGPQGPAG---KNGETGQGPPTGPGSDKGDGTG 633
 QY 228 -PKREGF--LPPPPPPPG-----GDRGRGPGGMRGGGGLMDRGPGGMFRGGRG 276
 DB 634 PPGPQGLQGLPGTSGPPGKPGEPKGEAGAPGIPGKKG---DSGAPGERGPPGAGG 690
 QY 277 DRGGRGGRGMDRGFGGRRGGPGPGPG---PLMEQMGRRGGRGGPG--KMDKGE 328
 DB 691 PPGP-RGGAG-PPGPEGKGAAGPPPPPSAGTPTGLQMPGERGGPGPGPKDKGE 745

RESULT 15

US-10-402-072A-6
 Sequence 6, Application US/10402072A
 Publication No. US20040018592A1
 GENERAL INFORMATION:
 APPLICANT: Bell, Marcum P.
 APPLICANT: Naff, Thomas B.
 APPLICANT: Polarek, James W.
 APPLICANT: Sealey, Todd W.
 TITLE OF INVENTION: BOVINE COLLAGENS AND GELATINS
 FILE REFERENCE: PP0402.2 CON
 CURRENT APPLICATION NUMBER: US/10/402,072A
 CURRENT FILING DATE: 2003-03-26
 PRIOR APPLICATION NUMBER: US 09/709,700
 PRIOR FILING DATE: 2000-11-10
 NUMBER OF SEQ ID NOS: 72
 SOFTWARE: PatentIn version 3.2
 SEQ ID NO 6
 LENGTH: 1466
 TYPE: PRT
 ORGANISM: Sus scrofa
 US-10-402-072A-6

Query Match 15.8%; Score 303; DB 15; Length 1466;
 Best Local Similarity 31.2%; Pred. No. 5.1e-13;
 Matches 130; Conservative 22; Mismatches 157; Indels 108; Gaps 21;

QY 1 GGRGMSAGEGRGFKRGGPMDEGPDLDGPPVDPEDSDNSAIYVQGLNDSTVTLDDLA 60
 DB 348 GAKGEVGPASSPGSSGAPGQGRGEPFGHAGAPPPPPSGNSPGKGEKMPAGIPGAP 407
 QY 61 DFFKQCGVVMKRRKTGQPMIHLYDK-----ETGK--PKGDATVSYE-----DPTAKAA 108
 DB 408 -----GLIGARPPPPGPTGTVGQGRGAAGEPGKNAKDDPPRGRGERGASFGIAGPK 461
 QY 109 VEMFDGKD-----FQSKLVSLARKKPPMNSMRG----- 138
 DB 462 GE--DGKDGSPGEBGANGLPGAAGERGVPPFRGPAGANGLPGEKGPDRGPGPAPRG 519
 QY 139 --GLPPREGMPPPLRGSGPGPGPG-----GPMGRMG-----GRGDDG-----G 178
 DB 520 VAEPPGIDGLPGGDLGRLGIPSGPGSDGKPPGSGQGETGRPPGSPGPPGQPGVWG 579
 QY 179 PPRGPRGSRGNPS-----GGNVQHRAGDMQCPNPGCGNQNFAMRTECNQCKA--- 227
 DB 580 FP--GPKNDGAPKNGERGGPGGPQGPAG---KNGETGQGPPTGPGSDKGDGTG 633
 QY 228 -PKREGF--LPPPPPPPG-----GDRGRGPGGMRGGGGLMDRGPGGMFRGGRG 276
 DB 634 PPGPQGLQGLPGTSGPPGKPGEPKGEAGAPGIPGKKG---DSGAPGERGPPGAGG 690
 QY 277 DRGGRGGRGMDRGFGGRRGGPGPGPG---PLMEQMGRRGGRGGPG--KMDKGE 328
 DB 691 PPGP-RGGAG-PPGPEGKGAAGPPPPPSAGTPTGLQMPGERGGPGPGPKDKGE 745

Search completed: February 18, 2005, 15:43:17
 Job time : 143 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 18, 2005, 15:22:27 / Search time 40 Seconds

(without alignments)
813.032 Million cell updates/sec

Title: US-10-791-017a-2_COPY_319_656

Perfect score: 1922

Sequence: 1 GGRGMSAGERGPFKPGG.....GGPGMDKGEHRRDRPV 338

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR 79: *
1: p1r1: *
2: p1r2: *
3: p1r3: *
4: p1r4: *

SUMMARIES

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No.	Score	Query Match	Length	DB ID	Description
1	1922	100.0	656	1 A49358	RNA-binding protei
2	1888.5	98.3	655	1 A55726	RNA-binding protei
3	878	45.7	526	1 S33799	RNA-binding protei
4	871	45.3	528	2 G02127	fus-1-like protein -
5	750	39.0	589	2 S71954	RNA/sedna-binding
6	623.5	32.4	404	2 S54729	RNA-binding protei
7	420.5	21.9	545	2 T15667	hypothetical prote
8	370.5	19.3	211	2 C96539	hypothetical prote
9	302	15.7	886	2 I50694	collagen alpha 1(I
10	301	15.7	1466	1 CGHUTL	collagen alpha 1(I
11	299.5	15.6	1049	1 CGBOTS	collagen alpha 1(I
12	298	15.5	310	2 I50696	collagen alpha 1(I
13	293	15.2	1464	2 S59856	collagen alpha 1(I
14	291	15.1	276	2 T33925	hypothetical prote
15	283.5	14.8	1042	1 CGCH1S	collagen alpha 1(I
16	278.5	14.5	326	2 A41732	heterogeneous ribo
17	277.5	14.4	1147	1 MMX1B	myosin heavy chain
18	275.5	14.3	1486	1 CGH2V	collagen alpha 2(V
19	271.5	14.1	960	1 S22315	snRNP-associated p
20	270	14.0	1419	2 JEO291	FB19 protein - hum
21	270	14.0	1487	2 A41182	collagen alpha 1(I
22	270	14.0	1487	2 B41182	collagen alpha 1(I
23	270	14.0	1492	2 A40333	collagen alpha 1(I
24	270	14.0	1497	2 I49607	procollagen type V
25	269.5	14.0	342	2 S14432	heterogeneous ribo
26	269	14.0	365	2 A26459	helix-desfibrillin
27	269	14.0	1486	1 B40333	collagen alpha 1(I
28	268.5	14.0	1758	2 T29350	hypothetical prote
29	265.5	13.8	1453	2 S21626	collagen alpha 1(I

30	265	13.8	352	2 T24279	hypothetical prote
31	265	13.8	1603	2 S23810	collagen alpha 1(X
32	264	13.7	435	2 T15143	hypothetical prote
33	263.5	13.7	1464	1 CGH01S	collagen alpha 1(I
34	263.5	13.7	1880	2 T18531	tractin - medicina
35	262	13.6	414	2 JN0866	nucleolar protein
36	262	13.6	3176	2 CGH03A	collagen alpha 3(V
37	261.5	13.6	671	1 CGRT1S	collagen alpha 1(I
38	261.5	13.6	779	1 CGB01S	collagen alpha 1(I
39	261.5	13.6	1418	2 T45467	collagen alpha 1(I
40	261.5	13.6	1806	1 CGB01E	collagen alpha 1(X
41	261	13.6	2944	2 A54849	collagen alpha 1(V
42	259.5	13.5	309	2 T19389	collagen alpha 1(V
43	259	13.5	346	1 S35500	hypothetical prote
44	259	13.5	385	2 S40778	heterogeneous ribo
45	258.5	13.4	1487	1 CGH06C	ribonucleoprotein
					collagen alpha 1(I

ALIGNMENTS

RESULT 1
A49358
RNA-binding protein EMS - human
C:Species: Homo sapiens (man)
C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 09-Jul-2004
C/Accession: A49358; S28257
R:Plougastel, B.; Zucman, J.; Peter, M.; Thomas, G.; Delattre, O.
Genomics 18, 609-615, 1993
A>Title: Genomic structure of the EMS gene and its relationship to EMSR1, a site of tum
A/Reference number: A49358; NCID:9410360; PMID:8307570
A/Accession: A49358
A/Status: nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA
A/Residues: 1-656 <RES>
A/Cross-references: UNIPROT:001844; EMBL:X72990; NID:9485838; PIDN:CAA51489.1; PID:9825.
R:Delattre, O.; Zucman, J.; Plougastel, B.; Desmarte, C.; Melot, T.; Peter, M.; Kovar, H
Nature 359, 162-165, 1992
A>Title: Gene fusion with an ETS DNA-binding domain caused by chromosome translocation
A/Reference number: S28257; NCID:92396239; PMID:1522903
A/Molecule type: mRNA
A/Residues: 1-656
A/Cross-references: EMBL:X66899; NID:9547565; PIDN:CAA47350.1; PID:931280
C/Genetics:
A/Gene: GDB:EMSR1
A/Cross-references: GDB:135984; OMIM:133450
A/Map position: 22q12.1-22q12.1
A/Intons: 5/1; 17/2; 34/3; 76/1; 138/2; 194/2; 265/1; 325/2; 338/1; 349/1; 388/3; 432/
A/Note: EMSR1 region is exons 7-10 of this gene, called EMS in reference A49358; this r
C/Superfamily: RNA-binding protein, EMS type; ribonucleoprotein repeat homology
C/Keywords: carcinogenesis; nucleus; RNA binding; tandem repeat
F:362-437/Domain: ribonucleoprotein repeat homology <RRM>

Query Match
Best Local Similarity 100.0%; Score 1922; DB 1; Length 656;
Matches 338; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	GGRGMSAGERGPFKPGGPMDEGPDLDGPPVDDDESDNSAIVVGGINDSVTLDDLA	60
DB	319	GGRGMSAGERGPFKPGGPMDEGPDLDGPPVDDDESDNSAIVVGGINDSVTLDDLA	378
QY	61	DFPKQGGVKKMKRTGQPMIHLYLDKETGPKPDATVSYEDPPTAKAAVWEPDGKDFQGS	120
DB	379	DFPKQGGVKKMKRTGQPMIHLYLDKETGPKPDATVSYEDPPTAKAAVWEPDGKDFQGS	438
QY	121	KAKVSLARKKKPPNNRSGCLPPREGGMPPLRGGGGGGGGGPGMRGSGKGDGCGPP	180
DB	439	KAKVSLARKKKPPNNRSGCLPPREGGMPPLRGGGGGGGGGPGMRGSGKGDGCGPP	498
QY	181	PKGPRGSRGNPSGGGVVORAGDMQCPNPGCGNPNAMTEGNOCCAPBEGFLPPFPFP	240
DB	499	PKGPRGSRGNPSGGGVVORAGDMQCPNPGCGNPNAMTEGNOCCAPBEGFLPPFPFP	558

Qy	24	PGGDRGSGCGMRRGGRGGGLMDRGCGGCFRGSRGDSDGFPGGGMMDRGCFGGGRGCGP	300
Dd	559	PGDDRGRGCGMRRGGRGGGLMDRGCGCFRGSRGDSDGFPGGGMMDRGCFGGGRGCGP	618
Qy	301	GGRPPLMEOMGSRGRCGRGSPGDMGGERORRRRPY	338
Dd	619	GGRPPLMEOMGSRGRCGRGSPGDMGGERORRRRPY	656

RESULT 2

RNA-binding protein Ews - mouse
C:Species: Mus musculus (house mouse)
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 09-Jul-2004
C:Accession: A55726, S45007
R:Ploegastel, B.; Mattei, M.G.; Thomas, G.; Delattre, O.
Genomics 23, 278-281, 1994
A>Title: Cloning and chromosome localization of the mouse Ews gene.
A:Reference: A55726; MUID:95130099; PMID:782090
A:Accession: A55726
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-655 <P/O>
A:Cross-references: UNIPROT:O61545; GB:X79233; NID:9486512; PIDN:CAA5815.1; PID:9486513
A:Note: authors translated the codon TCA for residue 116 as Thr, GCC for residue 123 as
C:Genetics:
A:Gene: Ews
C:Superfamily: RNA-binding protein, EWS type; ribonucleoprotein repeat homology
C:Keywords: carcinogenesis; nucleus; RNA binding; tandem repeat
f361-436/Domains: ribonucleoprotein repeat homology <RNM>

Query Match	98.3%	Score 1888.5;	DB 1;	Length 655;
Best Local Similarity	98.2%	Pred. No. 1.7e-111;		
Matches 332;	Conservative 4;	Mismatches 1;	Indels 1;	

QY 1 GGRGGMGSGERGGFNKPGGEMDEGDLTGAPVDEDEDSNSAIYQGLNDSVTLLDLA 60
 319 GGRGGLG-AGERGCFNKPGEEMDEGDLTLGPIDEDEDSNSAIYQGLNDSVTLLDLA 377
 Db

61 DFFKQCGVKNKRTGQPMIHIYLDKETGKPKGDATVSYEDPPTAKAAVEWFDGKDFQGS 120

Db 378 DFFKQCGVVKMKRTGQPMIHLYLDKETGKPKGDATVSYEDRPTAKAAVEWFDGKDFQGS 437

121 KLKVLARKKPPMNSMRGGLPPREGRGMPPILRGPGGGPGGPMGRMGGRGGDRGGFP 180

Db 438 KLKVS LARKKPPMNSMRGMPREGRGMPPLRGPGGPGGPGMGRMGGRGDRGGFP 497

181 PRGPRGSRGNPSGGNVQHRAGDWQCPNPGCGNQNFAMWTECNQCKAPKEGFLPPFPF 240

Db 498 PRGPRGSRGNPSGGGNVQHRAGDWQCPNPGCCGNQNFARTECNQCKAPKEGFLPPFP 557

241 PGDGRGGPGGMRGGGLMDRGCPGGMFRRGGGDRGGFRGGRCMDRGGFGGRRGGP 300

Db 558 PGDRGRGGPGMRRGGGLMDRGGPGMFRGGGRGGGFRGGRCMDRGGFGGGRGGP 617

301 GSPGPIMEQMGRRGGPGKMDKGEHRQERRDRPY 338

Db 618 GGP^PG^PLMEQMGRRGG^PGKMDKGEHRQERDRPY 655

RESULT 3

RNA-binding protein FUS, nuclear - human
N.Alternate names: RNA-binding protein TLS
C.Species: Homo sapiens (man)
C.Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 09-Jul-2004
C.Accession: S33799; S36157
R.Croizat, A.; Aman, P.; Mandahl, N.; Ron, D.
Nature 363, 640-644, 1993
A.Title: Fusion of C/EBP to a novel RNA-binding protein in human myxoid liposarcoma
A.Reference number: S33798; MUID:93286139; PMID:8510758
A.Accession: S33799

A;Molecule type: mRNA

A:Residues: 1-526 <CRO>
A:Cross-references: UNIPROT:P35637; GB:S62140; NID:G386156; PIDN:AAB27102.1; PID:G38615
1.Experimental source: 1105arcoma

A; Experimental source: [liposarcoma](#)
R; Rabbits, T.H.; Forster, A.; Larson, R.; Nathan, P.
Nature Genet 4 175-180 1993

A:Title: Fusion of the dominant negative transcription regulator CHOP with a novel gene
3-Reference number: 826157, PMID:93350637, PMID:7503811
Native gene: 7, 15-100, 1552

A/Reference number: S36157
A/Accession: S36157
A/Status: preliminary

n/Occasus: *preliminary*
 A/Molecule type: mRNA
 n/Postdates: 1-63 /S/ 66-526 (PAR)

A; Cr088-references: EMBL: X71428; NID: g393415; PIDN: CAA50559.1; PID: g4210363
A; Experimental source: J1nosarcoma

A:Gene: GDA.ETIS
C:Genetics:
n/labprimcatal bootcc; 229022200ma

A:Cross-references: GDB:136048; OMIM:137070
A:Map position: 16p11.2-16p11.2

n/map position: 1062112 1062112
c/function: RNA binding; probable n/a's a role in transcriptional regulation

Keywords: *carcinogenesis*; nucleus; RNA binding; tandem repeat C/Superfamily: RNA-binding protein, EMS type; ribonucleoprotein repeat homology

F;286-361/Domain: ribonucleoprotein repeat homology <RRM>

Query Match	45.7%;	Score 878;	DB 1;	Length 526;
Best local similarity	52.9%;	Pred. No. 3.8e-48;		

Matches 182; Conservative 36; Mismatches 54; Indels 72; Gaps 13;

QY 1 GGRGNGSAGERGCFNKPGGMDGPDLDGPVMDPDESDNSAIYVQGLNDSVTLDDLA 60

```

Db      249 GGGGMMGGS-DRGGENKFGRRDQGSRRD-----SEQDNSDNTTIFQGLGENVTIESVA 30

```

61 DFFKQGVWKNKRTGQPMIHIYLDKETGPKGDATVSYEDPTAKAAVENFDGKDFQS 12

Db 303 DYFKQIGI I KTNKKTGQPMINLYTDREYGLKGEATVSPDDPSAKAAIDMFDGKEFSGN 36

QY 121 KLVKSLARKKPPMNMSMRGGLPEPREGGNPPLRLGCGPGCGPGGPMGRMGGRGGDRGGCFP 18

Db 363 PIKVPATRRADFN--RGGGNGRGGRG---RGGPMGRGGGYGGG---GSGGGGGRGGPP 41

181 PRGPRGRGNPSSGGGVQHRAGDMQCPNPGCGGNQFAMWTECNOCKAKPEGGLPPPPPP 24

Db 412 SGG-----GGGGGGQRAAGDWKCNPPTCENNMFSSWRNECQCKAPKPKDG-----45

241 PGDGRGPGGMR--GGRGIMDRGPGMFRG--GRGDRGGFRGCR-GMDRGGFGGG 29

Db 455 PG-----GGPGGSHMGNGNYGDDRRGGGGRGGYDRGGYRGRRGGDRGGFRGGRRGGGDRGGF--- 50

296 RRGPGPPPLMEQMGRRGGRGGPPGKND-KGEHQERRDRPY 338

Db 507 -----GPGKMDSRGEHQDRERPPY 526

RESULT: 4
C02127

fus-like protein - human (fragment)
 C|Species: Homo sapiens (man)
 C|Date: 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change 09-Jul-2004
 C|Accession: G02127
 R|Itoh, K.; Kawase, M.
 submitted to the EMBL Data Library, September 1995
 A|Reference number: G09199
 A|Accession: G02127
 A|Status: preliminary; translated from GB/EMBL/DBJ
 A|Molecule type: mRNA
 A|Residues: 1-558 <IT>
 A|Cross-references: UNIPROT:013344; EMBL:U36561; NID:g1040969; PIDN:AAA9948.1; PID:g10
 C|Superfamily: RNA-binding protein, EMS type: ribonucleoprotein repeat homology
 F|289-364/Domain: ribonucleoprotein repeat homology <RKM>

Query Match	45.3%;	Score 871;	DB 2;	length 528;
Best Local Similarity	52.8%;	Pred. No. 1e-47;		
Matches 181;	Conservative 36;	Mismatches 54;	Indels 72;	Gaps 13

[illegible]

```

RESULT 5
S71954
RNA/ssDNA-binding protein TAFII68 - human
N/Alternate names: TATA-binding protein-associated factor TAFII68
C/Species: Homo sapiens (man)
C/Date: 19-Mar-1998 #sequence_revision 17-Apr-1998 #text_change 09-Jul-2004
C/Accession: S71954
R/Bertozzi, A.; Lutz, Y.; Heard, D.J.; Chambon, P.; Torá, L.
EMBL J. 15 5022-5031, 1996
A>Title: hTAFII68, a novel RNA/ssDNA-binding protein with homology to the pro-oncoprotein
A/Reference number: S71954; MUID:97045110; PMID:8890175
A/Molecule type: mRNA
A/Residues: 1-589 <BER>
A/Cross-references: UNIPROT:Q92804; EMBL:X98893; NID:g1628402; PIDN:CNA67398.1; PID:g1628402
A/Experimental source: HeLa cells
C/Gene(s):
A/Gene: taf68
C/Complex: up to 13 TATA-binding protein-protein-associated factors (TAFIs) form together with C/FUNCTION:
A>Description: required for activated transcription, binds RNA and single stranded DNA; C/Superfamily: RNA-binding protein, EMS type; ribonucleoprotein repeat homology
C/Keywords: RNA binding
F/232-307/bomaim: ribonucleoprotein repeat homology <RRM>

Query Match      39.0%, Score 750; DB 2; Length 589;
Best Local Similarity 39.3%; Pred. No. 4.5e-40;
Matches 175; Conservative 38; Mismatches 78; Indels 154; Gaps 15;

CY    2   GRGCM--GSAGERGCFNKPGSGPMDEGPDLIGPPVDDED--SDNSAIYVGGANDVTLLD 58
       ||| | | | | | | | | | | | | | | : | | | | | | | | | | | | | | | | 
Db    191 GRCGMTSSSGCDRGDFKNFGHGRDYGPRTD----ADBSDDSNDDNTITPVCGIAGEVSSTDQ 246
               :::::
OY    59  LADEFKCGGVNNKRTRTGQPMIHLYLDKEKGSKPKGDATVSSEDPPTAKAAVEWFDGDFQ 118
       ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | 
Db    247 VGEFFKQIGIKTKKKKTGKPMINLTDTDYDTKPKEALVSDDPSPAKAILDWFDGKEFH 306
               :::::
CY    119 GSKILVSLARKKRPMMNSNRGLPRREGCMPPLRGGPGGPGGCGGPMGWGRGGRGDRG 178
       : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | 
Db    307 GNLIKVFSTRPRPF--MRGC-----GSGGGRGRGGGYNG--RGG 342
               : | | | | | | | | | | | | | | | | | | | | | | | | | | | 

CY    179 PPPPGPSRKANBEGGAVCHRADMGCPNPGCCGNONFAMTTECNOCAPPEGFTLEPPF 238
       : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | 
Db    343 FQGQG---GDP-----KSGDWVCNPDSCGNMNFARINSNCQCNEPRFE-----DS 384
               : | | | | | | | | | | | | | | | | | | | | | | | | | | | 

CY    239 PPPGCD-----RGRGGPGGRGGRGGLM-----                261

```

Db	385	RPGSGDFRGRGVCGERGYGRGGRGDRGDRGDRSGGSGGDRSGGSGYG	444
Qy	262	-----DRGPGGMFRGGRGDRGGRF-----	281
Db	445	GDRSGGSGGDRGSGYGGDRGGGSGDRGGGSGGGRGGGRGGDRGGG	500
Qy	282	-----RGGRGMDRGGRFGGGR- RGGPGGPPGRLMEQMGGRGGRGGP-----	322
Db	505	YGDRGSGYGGDRGSGYGGDRSGYGGDRGSGSGYGGDRSGGSGYGGDRGGGSGYG	564
Qy	322	-----GKM-DKGRHGRDRDPY	338
Db	565	GDGSGYGGMGGRANDYRNDGRNRPY	589

```

RESULT 6
554729
RNA-binding protein cabeza - fruit fly (Drosophila melanogaster)
N/Alternate names: caz protein
C/Species: Drosophila melanogaster
C/date: 15-Jul-1995 #sequence_revision 01-Sep-1995 #text_change 16-Aug-2004
C/accession: S54729, S54728
R/Stolow, D.T.; Haynes, S.R.
submitted to the EMBL Data Library, October 1994
A/Description: Cabeza, a Drosophila gene encoding a novel RNA binding protein, shares h
A/Reference number: S54729
A/Accession: S54729
A/Molecule type: DNA
A/Residues: 1-404 <STO>
A/Cross-references: UNIPROT:Q27294; EMBL:L37083; NID:g567105; PIRN:AAC41563.1; PID:g567
R/Stolow, D.T.; Haynes, S.R.
Nucleic Acids Res. 23, 835-843, 1995
A/Title: Cabeza, a Drosophila gene encoding a novel RNA binding protein, shares homolog
A/Reference number: S54728; MUID:95223793; PMID:7708500
A/Accession: S54728
A/Status: nucleic acid sequence not shown
A/Molecule type: DNA
A/Residues: 118-203/273-310 <STW>
A/Cross-references: EMBL:L37083
C/Genetics:
A/Gene: cabeza
A/Cross-references: FlyBase:FBgn0011571
C/Superfamily: ribonucleoprotein repeat homology
C/Keywords: RNA binding
F/120-195/omam: ribonucleoprotein repeat homology <RMM>

Query Match      32.4%; Score 623.5; DB 2; Length 404;
Best Local Similarity 43.0%; Pred. No. 2.8e-32;
Matches 157; Conservative 32; Mismatches 125; Indels 51; Gaps 14;

QY      2 GRGGMGASGERGG-----FNK-PCGPMDEGDDLGPVPDDED--SDNSAIY 46
Db      63 GGGGGGSMNDRGNSYNGGASKDSYNGPGISGGGGGGGGGGGGGSGGNDMTQEDITF 122
QY      47 VQGLNDSVTLDDLADFFKQCGVVMNKRRTQPMIHILYDKETGPKGDATVSYEDPPTAK 106
Db      123 VSGMDPSTTEGDITRTHGALGIITIKDKRTMKPKIMLYKNKETGASKGSAVTTVDYDINAQ 182
QY      107 AAVVWFGKCPQSGKLVSLARKKPPMNSMRGCI-PPBEGRMPPPLGKPGCGPGCGPGM 166
Db      183 SAIEMFGRRDFNGNAIVTSILAQRNNWNKGGGGGGGGGGGGGFGRRGGGGGGGGGGGG 242
QY      167 GRM-----GARGDGRGFPFRGGRSGRGNPVGGSNVOHRAQDQCENPGCGNFAVRT 220
Db      243 GRFPRGGGGGGNGGGGGGGRIDRG--GGGGGGGGGGGQPRDGDYKNS--CANNTPFAWKN 298
QY      221 ECNQCKAPKEGFLPPFPPEPGDGRGCGPGMRGRCGLMDRGGPGCGMFRGRCGDRGG 280
Db      299 EGNRCKTPK-----GDDESSGGGG--GGYGG-----GGGGGGY--DRGNDRGS 339
QY      281 PRGG-RGMDRG--FGGRRGGPGGPPGLMEQNGRRGRGGRG--KMDKGEHQE--R 333
Db      340 GGGGYNHRDRGNSQGGGGGGGGGGGGYSRFDNNNGGRRGGGGGGGGGNRRDGGPRMNDGGM 399

```


QY 334 RDRPY 338
DB 400 RSRPY 404

RESULT 7

hypothetical protein C27H5.3 - *Caenorhabditis elegans*

C/Spectes: *Caenorhabditis elegans*

C/Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #ext_change 09-Jul-2004

C/Accession: T15667

R/Author: A.

submitted to the EMBL Data Library, July 1995

A/Description: The sequence of *C. elegans* cosmid C27H5.

A/Reference number: Z18386

A/Accession: T15667

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-545 <PAU>

A/Cross-references: UNIPROT:Q18265; EMBL:U14635; NID:9540265; P1D:9540269; P1DN:AA046657

A/Experimental source: strain Bristol N2

C/Genetics:

A/Genes: C27H5.3

A/Intons: 36/2; 52/1; 87/3; 183/1; 202/2; 262/1; 308/2; 511/3

C/Superfamily: RNA-binding protein, EMS type; ribonucleoprotein repeat homology

Query Match 21.3%; Score 420.5; DB 2; Length 545;

Best Local Similarity 30.6%; Pred. No. 2e-19;

Matches 136; Conservative 29; Mismatches 95; Indels 184; Gaps 19;

QY 1 GGRGGM-----GSAGRGGF-----NKGPG 20

DB 180 GGRGGYDGERRRGSRMDGNSDRGGPGRGGYQGIYVNCIFRTTSLFLKKNQY 239

QY 21 PMD-----EGPDLDGPPV-----DPDEDSD-----NS 43

DB 240 PTFEITIEIQOKTLENYKNTCSYRGPRRD--GPPSGGSGGGAASGNRBERGSDRYLKE 298

QY 44 AIYVQGLNDVTLDDLADFPKQGVVNMKRTQPMHIYLDKTEGPKADATVSYEDP 103

DB 299 TVFVQGISITANAYADVSTGDIAXKND--GPKIKYTDNTGPKCECMITTVDA 356

QY 104 TAAAVEMFPGKDFQ--SLTKYSLARKKPPMNSMRGSLPRRGRGPPRLRGSPG 160

DB 357 AAQQAATMTNGQFPFGSSSMISLAKFRADACGRG-----KGRGSPG 403

QY 161 GPGSPMGRMGGRGDRGPPRGRGSRGNPSGGGVQHRAGDMQCPNPGCNGNPFAMRT 220

DB 404 GRGPMGRGGRGFGGDRGY---GGGGGRG----- 429

QY 221 ECHQCAKPKREGFLPPPPPPGDRGRGSGMRGSGGLMDRGSGMFRG--RGSDRG 279

DB 430 -----GPDGSGGGGSGFRG-----DRGFRGSGGSGFRGSDRG 463

QY 280 GFRGSG--RGMDRGSGGSGRGRGSGPPEPL-----MEQMG-----RR----- 315

DB 464 GFRGSDRGSGRGRGR--GVGGGANMMQRKNDPCECGSNRFAFRRCNQQCAFRP 521

QY 316 -GGRGSPGRKMDKEHNRDRRPY 338

DB 522 DGGSGGGGGRGGRGSGPGRGDRRPY 545

RESULT 8

C96539

hypothetical protein F1413.10 [imported] - *Arabidopsis thaliana*

C/Spectes: *Arabidopsis thaliana* (mouse-ear creas)

C/Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #ext_change 09-Jul-2004

C/Accession: C96539

R/Author: A.; Eckert, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,

Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;

ansen, N.F.; Hughes, B.; Hutzar, L.

Nature 408, 816-820, 2000

A/Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.

C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luce, J.S.; Malt, R.; Marzall,

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A/Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,

ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A/Title: Sequence and analysis of chromosome 1 of the plant *Arabidopsis*.

A/Reference number: A66141; MUID:21016719; PMID:11130712

A/Accession: C96539

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-211 <STO>

A/Cross-references: UNIPROT:Q95X47; GB:AE005173; NID:95734788; P1DN:AA050053.1; GSPDB:G

C/Genetics:

A/Genes: F1413.10

A/Map position: 1

Query Match 19.3%; Score 370.5; DB 2; Length 211;

Best Local Similarity 36.7%; Pred. No. 1.2e-16;

Matches 84; Conservative 21; Mismatches 91; Indels 33; Gaps 3;

QY 42 NSAIYQGLNDVTLDDLADFPKQGVVNMKRTQPMHIYLDKTEGPKADATVSYED 101

DB 7 NSVYVSNLPGLTNDMLADYFTIGLKRDRKTRPKWLYRDXETDEPKGDATVYED 66

QY 102 PTAAXAVEMFPGKDFQ--SLTKYSLARKKPPMNSMRGSLPRRGRGPPRLRGSPG 161

DB 67 PTAAXAVEMFPGKDFQ--SLTKYSLARKKPPMNSMRGSLPRRGRGPPRLRGSPG 114

QY 162 PGSPMGRMGGRGDRGPPRGRGSRGNPSGGGVQHRAGDMQCPNPGCNGNPFAMRT 221

DB 115 TTGAGR-----GRGQDSAKXWQDDGDMCENSTCTVNFAPRGV 156

QY 222 CHQCAKPKREGFLPPPPPPGDRGRGSGMRGSGGLMDRGSGMFRG--RGSDRG 270

DB 157 CRRGCTARPAAGASGSM--GAGRGGRGSGADGAPGKSPGAPGLP 202

RESULT 9

collagen alpha 1(III) chain - chicken (fragment)

C/Spectes: *Gallus gallus* (chicken)

C/Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #ext_change 09-Jul-2004

C/Accession: I50694

R/Author: H.D.; Niu, Z.; Adams, S.L.

A/Title: An alternative transcript of the chick type III collagen gene that does not en

A/Reference number: A54041; MUID:94266842; PMID:8206952

A/Accession: I50694

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: mRNA

A/Residues: 1-886 <NAH>

A/Cross-references: UNIPROT:P12105; EMBL:U07973; NID:9520454; P1DN:AAA83407.1; P1D:95337

C/Genetics:

A/Genes: COL3A1

C/Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology

P;30-90/Domain: von Willebrand factor type C repeat homology <VWC>

Query Match 15.7%; Score 302; DB 2; Length 886;

Best Local Similarity 30.8%; Pred. No. 8.8e-12;

Matches 114; Conservative 20; Mismatches 104; Indels 132; Gaps 22;

QY 1 GGRGSGSAGEGGRGFRKPG--GPMDEGPDLDL-----GPPYDPEDSDNSAIYVQGLN 51

DB 584 GNEGAGKNGERKPGPPTGPAGKNGVGLGPPGPAAGPADRGEPGSGPGLQGLP 643

QY 52 DSVTLDDLADFPKQGVVNMKRTQPMHIYLDKTEGPKADATVSYEDPTAAXAVEM 111

DB 644 GGP-----GPAENKRGEP-----GPKD----- 663

QY 112 PDKDFQSGSKLTKVSLARKKPPMNSMRGSLPRRGRGPPRLRGSPG-----GPG 163

DB 664 IGGPGRGPK-----GENGIRGERGPGPGPPTGARGGPGAGSGAGKGP 709

QY 164 GPMGMRGSG-----GDRGGRPRGSRGSRGNBSCGNGHRAADNCPNPGC-----211
DB 710 GPPGAPGSGTGLPGLQGMPSRGASGSPGPKGDGPBGK-----ADGL---FGARGER 760
QY 212 GNQNFAMRTECNQCKAPKPEGFLPPPPPP-PCGDRGRGPGG---RGRGGLMDRG---264
DB 761 GN-----VGPIGPPOPPAPPPDKETGTGAGAPGAPGASRGSGRGERGEG 803
QY 265 --GPGMFRG--GRGDRGFRGGRGMD--RGGFG-----GARRGPG--GPPGLMEQMG 312
DB 804 LPPGAP-PPGAPGQNGBPBG-KGBRGPGLRGAGPPGAGAPGCGPPGAPPPPP--GGVK 859
QY 313 GRRGGRGRC 322
DB 860 GERGSPGPG 869

RESULT 10
CGHUTL
collagen alpha 1(III) chain precursor - human
N/Alternate names: procollagen alpha 1(III) chain
C/Species: Homo sapiens (man)
C/Date: 24-Apr-1984 #sequence revision 01-Sep-1995 #text change 09-Jul-2004
C/Accession: S05272, S04642, PE0011, S01726, S04887, A90399, A94562, I51868, S59511, A90
R/Prockop, D.J.
Submitted to the EMBL Data Library, February 1989
A/Reference number: S05272
A/Accession: S05272
A/Status: preliminary
A/Molecule type: mRNA
A/Residues: 1-1240, 'V', 1242-1466 <PRC>
A/Cross-references: UNIPROT:P02461; EMBL:X14420, NID:G30057, PIDN:CAA32583.1; PID:G30058
R/Ala-Kokko, L.; Kontusheari, S.; Baldwin, C.T.; Kivianiemi, H.; Prockop, D.J.
Biochem. J. 260, 509-516, 1989
A/Title: Structure of cDNA clones coding for the entire prepro-alpha1(III) chain of human
collagen.
A/Reference number: S04642; NID:89350838; PMID:2764886
A/Accession: S04642
A/Molecule type: mRNA
A/Residues: 1-1196 <ALA>
A/Cross-references: GB:M26939, NID:G180813, PIDN:AAA52040.1; PID:G180814
R/Toman, P.D.; Ricca, G.A.; de Crombrughe, B.
Nucleic Acids Res. 16, 7201, 1988
A/Title: Nucleotide sequence of a cDNA coding for the amino-terminal region of human pro
collagen.
A/Reference number: S01726; NID:88303360; PMID:3405773
A/Accession: S01726
A/Molecule type: mRNA
A/Residues: 1-170 <OM>
A/Cross-references: GB:X07240, NID:G30060, PIDN:CAA30229.1; PID:G30061
A/Note: the authors translated the codon CAG for residue 154 as His
R/Janczko, R.A.; Ramirez, F.
Nucleic Acids Res. 17, 674, 1989
A/Title: Nucleotide and amino acid sequences of the entire human alpha-1 (III) collagen.
A/Reference number: S04887; NID:89386015; PMID:2780304
A/Accession: S04887
A/Molecule type: mRNA
A/Residues: 149-163, 'G', 164-240, 'D', 242-471, 'D', 473-487, 'L', 489, 'S', 491-613, 'Y', 615-634,
A/Cross-references: EMBL:X15332, NID:G29545, PIDN:CAA33367.1; PID:G30045
A/Note: the authors' translation of residues 905-932 is inconsistent with the nucleotide
R/Seyer, J.M.; Kang, A.H.
Biochemistry 16, 1158-1164, 1977
A/Title: Covalent structure of collagen: amino acid sequence of cyanogen bromide peptide
A/Reference number: A90399; NID:77134724; PMID:557335

A/Accession: A90399
A/Molecule type: protein
A/Residues: 'V', 169-225, 229-232, 'P', 234-292, 'D', 294-398 <SEY1>
A/Experimental source: liver
A/Note: sequence corrected by A94562; attachment of 2-O-alpha-D-glucosyl-O-beta-D-galac
tose.
R/Seyer, J.M.
Submitted to the Atlas, December 1977
A/Reference number: A94562
A/Accession: A94562
A/Molecule type: protein
A/Residues: 'V', 169-225, 229-277, 'A', 279-292, 'D', 294, 'S', 296-398 <SEY2>
A/Experimental source: liver
A/Note: author submitted corrections to A90399
R/Milewicz, D.M.; Wiltz, A.M.; Smith, A.C.; Manchester, D.K.; Waldstein, G.; Byers, P.H.
Am. J. Hum. Genet. 53, 62-70, 1993
A/Title: Parental somatic and germ-line mosaicism for a multiection deletion with unusual
fingerprint.
A/Reference number: I51868; NID:93304430; PMID:8317500
A/Accession: I51868
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 186-194 <MTL>
A/Cross-references: GB:S62925, NID:G386425, PIDN:AAJ3937.1; PID:G4261637
R/Chiodo, A.A.; Silience, D.O.; Cole, W.G.; Bateman, J.F.
Biochem. J. 311, 939-943, 1995
A/Title: Abnormal type IIT collagen produced by an exon-17-skipping mutation of the COL
1A1 gene.
A/Reference number: S59511; NID:96067614; PMID:7487954
A/Accession: S59511
A/Molecule type: mRNA
A/Residues: 302-423 <CHI>
A/Cross-references: GB:S79877, NID:G1195576, PIDN:AA35615.1; PID:G1195577
R/Seyer, J.M.; Kang, A.H.
Biochemistry 17, 3404-3411, 1978
A/Title: Covalent structure of collagen: amino acid sequence of five consecutive CNBr p
eptides.
A/Reference number: A90414; NID:79000343; PMID:687591
A/Accession: A90414
A/Molecule type: protein
A/Residues: 399-675, 'N', 677-727 <SEY3>
A/Experimental source: liver
R/Lea, B.; Vitale, B.; Superli-Furga, A.; Steinmann, B.; Ramirez, F.
J. Biol. Chem. 266, 5256-5259, 1991
A/Title: G to T transversion at position +5 of a splice donor site causes skipping of ti
ssue-specific exon.
A/Reference number: I55349; NID:91161621; PMID:1672129
A/Accession: I55349
A/Status: translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 537-605 <LBS>
A/Cross-references: GB:M59312, NID:G180815, PIDN:AAA52041.1; PID:G180816
R/Seyer, J.M.; Mainardi, C.; Kang, A.H.
Biochemistry 19, 1583-1589, 1980
A/Title: Covalent structure of collagen: amino acid sequence of alpha1 (III)-CDS from t
he rat.
A/Reference number: A90438; NID:80198282; PMID:6246925
A/Accession: A90438
A/Molecule type: protein
A/Residues: 728-895, 'A', 897-964 <SEY4>
A/Experimental source: liver
R/Cole, W.G.; Chiodo, A.A.; Lemande, S.R.; Janczko, R.; Ramirez, F.; Dahl, H.H.M.; Cha
u, J. Biol. Chem. 265, 17070-17077, 1990
A/Title: A base substitution at a splice site in the COL3A1 gene causes exon skipping a
nd a frameshift.
A/Reference number: A38303; NID:91009133; PMID:2145268
A/Accession: A38303
A/Molecule type: mRNA
A/Residues: 861-1015 <COL>
A/Cross-references: GB:J05617, GB:M55603, GB:M59227, NID:G180878, PIDN:AA59383.1; PID:
A/Note: a mutant sequence with 942-977 spliced out from a patient with Ehlers-Danlos sy
ndrome.
R/Hancock, B.S.; Dalglish, R.
Nucleic Acids Res. 16, 2337, 1988
A/Title: Human pro alpha1(III) collagen: cDNA sequence for the 3' end.
A/Reference number: S02119; NID:88189827; PMID:3357782
A/Accession: S02119
A/Status: translation not shown
A/Molecule type: mRNA
A/Residues: 950-1018, 'Y', 1020-1183, 'S', 1185-1466 <MAN>

A.Cross-references: EMBL:X06700; NID:g30053; PIDN:CAA29886.1; PID:g30054
B.Seyer, J.M.; Kang, A.H.
Biochemistry 20, 2621-2627, 1991
A>Title: Covalent structure of collagen: amino acid sequence of alpha1 (III)-C9 from cy
A.Reference number: A90446; MUID:81208139; PMID:7016180
A.Accession: A90446
A.Molecule type: protein
A.Residues: 965-979, 'N', 981-984, 'PS', 987, 'QN', 990-1096, 'P', 1098-1152, 'AT', 1155, 'S', 1157-
A.Experimental source: liver
R.IoIdi, H.R.; Brinker, J.M.; May, M.; Pihlajaniemi, T.; Morrow, S.; Rosenbloom, J.; Mye
Nucleic Acids Res. 12, 9383-9394, 1984
A>Title: Molecular cloning and carboxyl-1-propeptide analysis of human type III procollage
A.Reference number: A93551; MUID:85087944; PMID:6096827
A.Accession: A93551
A.Molecule type: mRNA
A.Residues: 1065-1155, 'P', 1157-1466 <LOI>
A.Cross-references: EMBL:X01655; EMBL:X01742, NID:g29584; PIDN:CAA25821.1
R.Miskuln, M.; Dalgleish, R.; Kluge-Beckerman, B.; Rennard, S.I.; Tolstoshev, P.; Brand
Biochemistry 25, 1408-1413, 1986
A>Title: Human type III collagen gene expression is coordinately modulated with the type
A.Reference number: 152393; MUID:86187804; PMID:3754462
A.Accession: 152393
A.Status: translated from GB/EMBL/DBD
A.Molecule type: mRNA
A.Residues: 1161-1200 <MIS>
A.Cross-references: GB:M13146; NID:g180415; PIDN:AA52003.1; PID:g180416
R.Remann, B.S.; Camitzaro, L.A.; Seyer, J.M.; Myers, J.C.
Proc. Natl. Acad. Sci. U.S.A. 82, 3385-3389, 1985
A>Title: Human alpha 1(III) and alpha 2(V) procollagen genes are located on the long arm
A.Reference number: 159025; MUID:85216505; PMID:3858826
A.Accession: 179359
A.Status: translated from GB/EMBL/DBD
A.Molecule type: mRNA
A.Residues: 1165-1196 <EWA>
A.Cross-references: GB:M11134; NID:g180417; PIDN:AA52004.1; PID:g180418
R.Chu, M.L.; Weil, D.; de Wet, W.; Bernard, M.; Sipola, M.; Ramirez, F.
J. Biol. Chem. 260, 4357-4363, 1985
A>Title: Isolation of cDNA and genomic clones encoding human pro-alpha1(III) collagen. H
A.Reference number: A92516; MUID:85157600; PMID:2579949
A.Accession: A92516
A.Molecule type: DNA
A.Residues: 1176-1240, 'V', 1242-1356, 'P', 1358-1466 <CHU>
A.Cross-references: GB:M10615; GB:M10793; GB:M10794; GB:M10795; GB:M10796; GB:M10797; GB
A.Experimental source: liver
A.Note: the authors translated the codon TTC for residue 1057 as Tyr; the codons given f
action
C.Comment: Prolines and lysines at the third position of the tripeptide repeating unit
3-hydroxylated. About 15% of the lysines are 3-hydroxylated and some are subsequently C
C.Genetics:
A.Gene: GDB:COL3A1
A.Cross-references: GDB:118729; OMIM:120180
A.Map position: 2q31-2q31
A.Introns: 27/1; 94/3; 111/3; 149/3; 176/3; 554/3; 587/3; 1175/3; 1275/1; 1337/3; 1418/3
A.Note: the 1st of introns is incomplete; defects in this gene can result in Ehlers-Dan
C.Complex: type III collagen is a homotrimer of monomers initially linked by disulfide b
er of their length, is formed with desmosine cross-links made from lysine and allysine
C.Function:
A.Description: structural component of extracellular fibrous polymer that maintains inte
C.Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
C.Keywords: coiled coil; Ehlers-Danlos syndrome; extracellular matrix; glycoprotein; hyd
P.1-22/Domains: signal sequence #status predicted <SIG>
P.124-153/Domains: amino-terminal propeptide #status predicted <PRO>
P.31-91/Domains: von Willebrand factor type C repeat homology <WVC>
P.154-1221/Product: collagen alpha 1(III) chain #status predicted <MAR>
P.154-167/Region: amino-terminal nonhelical telopeptide
P.168-1166/Region: helical
P.1091-1093/Region: cell attachment (R-G-D) motif
P.1197-1221/Region: carboxyl-terminal nonhelical telopeptide
P.1232-1466/Domains: carboxyl-terminal propeptide #status predicted <CPR>
P.1232-1466/Domains: fibrillar collagen carboxyl-terminal homology <FCC>
P.124/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status predicted
P.153-154/Cleavage site: Pro-Gln (procollagen N-endorphinase) #status predicted
P.154/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status predicted

Query Match 15.7%; Score 301; DB 1; Length 1466;
Best Local Similarity 30.5%; Pred. No. 1,66-11;
Matches 116; Conservative 15; Mismatches 125; Indels 124; Gaps 18;

F.161,1212/Modified site: allylsine (Lys) #status predicted
F.161,284,860,977,1106/Modified site: 5-hydroxylysine (Lys) #status experimental
F.161/Binding site: carbohydrate (Lys) (covalent) #status experimental
F.158,1094/Modified site: 5-hydroxylysine (Lys) (partial) #status experimental
F.148,949/Cleavage site: Gly-Ile (collagenase) #status experimental
F.1106/Binding site: carbohydrate (Lys) (covalent) #status predicted

QY 1 GGRGMSAGRGGGFNKPGMGPDLDIGPPVDPEDSNSAIVYGINDSVTLDDL A 60
DB :||| ||| |
438 GAKEPGRGRGEAGIPGVDAKEDKDGSGPEG-----GAN----- 476
QY 61 DFFNQCVVKNKKTKGTQPMHIIYDKETGKRKGATVSIEDPTPKAAVEFDKD FGS 120
DB :||| ||| |
477 -----GLPGAAGERGA PGR-----GPAGNNG PGEGKPAGE----- 508
QY 121 KLTAKSLARKRPNNMNM--GLPREGRGMPPRLRGGPGCGG-----GPMGRWG--G 171
DB :||| ||| |
509 -----RGAPDPAEPGRAMAEPRDDVPGGPGFKRMGPSFGGSGSKGPPPSQS ESE 561
QY 172 RGDGRGGFPKPGSRG-----NPSGGNVORHAGMQCPNPGCQNPFAMRTCN QCK 226
DB RPGRPG-----PGRRGQGQGWMPFPGPKNDAGA PKNGRGGPG--GPGQG----- 606
QY 227 APKEBGLPFPFP-----PGDRGRGGPGCKRG-----GRGLMDR----- 263
DB 607 PPKNGTGP GP GP GT PG PG DK DT GP PP GG I Q GL PT GT GP PR ENG K PG EP PK DAG 666
QY 264 --GGPGCMFRGSGRDG-----GPRGCGMDRGGFGGGRGPGGPPG----PLME 309
DB 667 APGAPGKCDAGAPERGGPEGLAGAPGLRGAG--PPREGGKLAAGPPGPMAAGTPQL Q 725

QY 310 QMGRRGGRGGP--KMDXGE 328
DB 726 GMPERGGLSPGPKDXGE 745

RESULT 11
CSB07S
collagen alpha 1(III) chain - bovine
C.Species: Bos primigenius taurus (cattle)
C.Date: 04-Dec-1986 #sequence revision 04-Dec-1986 #text change 09-Jul-2004
C.Accession: A02862; A38001; A38002; A38003; A38004; A38005; S71946
R.Fietzek, P.P.; Allmann, H.; Rauterberg, J.; Henkel, W.; Wachter, B.; Kuehn, K.
Hoppe-Seyler's Z. Physiol. Chem. 360, 809-820, 1979
A.Title: The covalent structure of calf skin type III collagen.
A.Reference number: A02862; MUID:80026026; PMID:488906
A.Accession: A02862
A.Molecule type: Protein
A.Residues: 1-242 <FI>
A.Cross-references: UNIPROT:P04258
R.Dewes, H.; Fietzek, P.P.; Kuehn, K.
Hoppe-Seyler's Z. Physiol. Chem. 360, 821-832, 1979
A.Title: The covalent structure of calf skin type III collagen. II. The amino acid sequ
A.Reference number: A38001; MUID:80026027; PMID:488907
A.Accession: A38001
A.Molecule type: Protein
A.Residues: 243-422 <DEMI>
R.Bentz, H.; Fietzek, P.P.; Kuehn, K.
Hoppe-Seyler's Z. Physiol. Chem. 360, 833-840, 1979
A.Title: The covalent structure of calf skin type III collagen. III. The amino acid sequ
A.Reference number: A38002; MUID:80026028; PMID:488908
A.Accession: A38002
A.Molecule type: Protein
A.Residues: 423-571 <BEN>
R.Liang, H.; Glanville, R.W.; Fietzek, P.P.; Kuehn, K.
Hoppe-Seyler's Z. Physiol. Chem. 360, 841-850, 1979
A.Title: The covalent structure of calf skin type III collagen. IV. The amino acid sequ
A.Reference number: A38003; MUID:80026029; PMID:488909
A.Accession: A38003

A/Molecule type: protein
 A/Residues: 572-808 <LAN>
 R/Dewes, H.; Fietzek, P.P.; Kuehn, K.
 Hoppe-Seyler's Z. Physiol. Chem. 360, 851-860, 1979
 A/Title: The covalent structure of calf skin type III collagen. V. The amino acid sequence
 A/Reference number: A38004; MUID:80026030; PMID:488910
 A/Accession: A38004
 A/Molecule type: protein
 A/Residues: 809-947 <DEM2>
 R/Allmann, H.; Fietzek, P.P.; Glanville, R.W.; Kuehn, K.
 Hoppe-Seyler's Z. Physiol. Chem. 360, 861-868, 1979
 A/Title: The covalent structure of calf skin type III collagen. VI. The amino acid sequence
 A/Reference number: A38005; MUID:80026031; PMID:488911
 A/Accession: A38005
 A/Molecule type: protein
 A/Residues: 948-1049 <ALL>
 R/Experimental source: skin
 R/Henkel, W.
 Biochem. J. 318, 497-503, 1996
 A/Title: Cross-link analysis of the C-telopeptide domain from type III collagen.
 A/Reference number: S71946; MUID:96404897; PMID:8809038
 A/Accession: S71946
 A/Molecule type: protein
 A/Residues: 87-106/1017-1039/1037-1049 <HEN>
 C/Comment: Prolines at the third position of the tripeptide repeating unit (G-X-Y) are H
 C/Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
 C/Keywords: cold coil, extracellular matrix, glycoprotein, hydroxylysine, hydroxyproline
 F/1-1049/Product: collagen alpha 1(I) chain #status experimental <CAB>
 F/1-14/Region: amino-terminal nonhelical telopeptide
 F/15-1040/Region: helical
 F/587-589/Region: cell attachment (R-G-D) motif
 F/752-754/Region: cell attachment (R-G-D) motif
 F/875-877/Region: cell attachment (R-G-D) motif
 F/878-880/Region: cell attachment (R-G-D) motif
 F/935-937/Region: cell attachment (R-G-D) motif
 F/1041-1049/Region: carboxyl-terminal nonhelical telopeptide
 F/95-107, 119, 938, 950/Modified site: 5-hydroxylysine (lys) #status experimental
 F/107/Binding site: carboxylate (lys) (covalent) #status experimental
 F/1040, 1041/Dissulfide bonds: interchain #status predicted

Query Match 15.6%; Score 299.5; DB 1; Length 1049;
 Best Local Similarity 32.2%; Pred. No. 1.5e-11;
 Matches 133; Conservative 14; Mismatches 121; Indels 145; Gaps 25;

QY 1 GGRGSGSAGERGFGKPGPMDEGPDLDGPPVDPEDSDNSAIYVQGLN 60
 DB 237 GKGGEWGPAGIPGA---PGLIGARPP---GPP-----GTN----- 266

QY 61 DFFKQGVVAKMKTQPMHIYLDKXETGK--EKGDATVSYE---DPTAKAWEWD 113
 DB 267 ---GVPGQGRGAGGP-----GKNGAKGDPGPRGERGEAGSPGIAPKGE--D 308

QY 114 GKD-----FQSKLKVSLARKKTPMNSMRGGLPRREGGMP--- 149
 DB 309 GKDGSFGEAGNGLPQAGRGVPGFRGPAGANGLGKGGPPGD--RGGRPAGPRVAGE 367

QY 150 -----PPLRGPGPGPGPGP---GPMGRMG---GRGDRG---GFPRP 182
 DB 368 PGRNGLPGRGLRGIRGSPGPGPSNKKPGRSGQGTGRGPRGSPGPGQPGVWGF-- 425

QY 183 GPRGRGNNS-----GGNVQHRADWQCPNPGCGNQNPAMRTENQCKA---PKP 230
 DB 426 GPRKNGAGRGKNGERGPGPGPGQPGAG---KNGSTGPGPGPTGPGSGDKDTPGPP 481

QY 231 EGF--LPPFPFPPG-----GDRGRGPGMRGGRGGLMRGSGGMRFRGRGGRD 280
 DB 482 QGLQGLPTSGSPGPGNGKRGEEPRKGEAPGIPGSGK---DSGAGERGPPAGGPPGP 538

QY 281 PRGRGMDRGGRGGRGPGPGPG---PLMEQWGRGRGGRGPG--KMDKGE 328
 DB 539 -RGAG--PGPRGKGAGAPPPGPRGAGTFLQGMGRGGRGPGPGPKDKGE 589

RESULT 12
 150696
 collagen alpha 1(III) chain - chicken (fragment)
 C/Species: Gallus gallus (chicken)
 C/Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 09-Jul-2004
 C/Accession: 150696
 R/Nah, H.D.; Niu, Z.; Adams, S.L.
 J. Biol. Chem. 269, 16443-16448, 1994
 A/Title: An alternative transcript of the chick type III collagen gene that does not en
 A/Reference number: A54041; MUID:9426842; PMID:8206952
 A/Accession: 150696
 A/Status: preliminary; translated from GB/EMBL/DBS
 A/Molecule type: mRNA
 A/Residues: 1-310 <NAH>
 A/Cross-references: UNIPROT:Q90612; EMBL:U07974; NID:9520456; PIDN:AAA83409.1; PID:G537
 A/Genetic: COL3A1
 C/Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology

Query Match 15.5%; Score 298; DB 2; Length 310;
 Best Local Similarity 30.2%; Pred. No. 6e-12;
 Matches 111; Conservative 23; Mismatches 106; Indels 128; Gaps 21;

QY 1 GGRGSGSAGERGFGKPG--GPMDEGPDLD-----GPPVDPEDSDNSAIYVQGLN 51
 DB 8 GMEGAGKNGERGPGRPGPRPGAGKNGVGLPGRPPAGPADNRGEPSPSGPGLQGLP 67

QY 52 DSVTLDDLDFKQGVVAKMKTQPMHIYLDKXETGKPKGDATVSYEDPTAKAWEWD 111
 DB 68 GGP-----GPAGENGKPGEP-----GREGD----- 87

QY 112 FPKDFQSKLKVSLARKKTPMNSMRGGLPRREGGMPPLRLRGPGPGP-----GPG 163
 DB 88 IGGRGPRPK-----GENGLPGERGAGCPGPGTCARGRGPRGSEGAKGPR 133

QY 164 GPMGRMGGRG-----GDRGFPFRGSRGSRNPSSGG--NVQHRADWQCPNPGCGN 213
 DB 134 GPRGAGGTGLPQLQGMPEGRGSGSPGPKGKEPGKAGADLPGERGE-----RGN 186

QY 214 QNPAMRTENQCKAPRBEFLPPPPPP--PGDGRGRGPGGM---RGGRGLMDRG--- 264
 DB 187 -----VPIGPAGPGPGGKGETGTPAGAPRASPGRGPGRGGQGLP 229

QY 265 GPGMFRG--GRGDRGFRGGRGMD--RGCFG---GGRRGPG--GPPPLMEQWGR 314
 DB 230 GPAG--PRGAPQNGERGG--KGERGPRGLRGAGPRGAPGPGGPRAGPRFP--GQVKGE 285

QY 315 RGRGSGPG 322
 DB 286 RGSPPGPG 293

RESULT 13
 S59856
 collagen alpha 1(III) chain precursor - mouse
 C/Species: Mus musculus (house mouse)
 C/Date: 10-Apr-1996 #sequence_revision 19-Apr-1996 #text_change 03-Jul-2004
 C/Accession: S59856; S62120; S16373
 R/Toman, P.D.; de Crombrughe, B.
 Gene 147, 161-168, 1994
 A/Title: The mouse type-III procollagen-encoding gene: genomic cloning and complete DNA
 A/Reference number: S59856; MUID:95011609; PMID:7926795
 A/Accession: S59856
 A/Molecule type: DNA
 A/Residues: 1-1464 <TOM>
 A/Cross-references: UNIPROT:P08121; EMBL:X52046
 R/Toman, D.
 submitted to the EMBL Data Library, November 1994
 A/Reference number: S62120
 A/Accession: S62120
 A/Molecule type: DNA

A/Residues: 1-866, 'G', 868-1464 <NO>
 A/Cross-references: EMBL:X52046; NID:G575321; PIDN:CAA56279.1; PID:G575322
 R/Reference: M.; Tomlin, D.; de Crombrughe, B.; Vuorio, E.
 Biochim. Biophys. Acta 1089, 241-243, 1991
 A/Title: Specific hybridization probes for mouse type I, II, III and IX collagen mRNAs.
 A/Reference number: S16176; MUID:91274355; PMID:2054384
 A/Accession: S16373
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1442-1464 <MET>
 A/Cross-references: EMBL:X57983; NID:G50476; PIDN:CAA1048.1; PID:G50477
 C/Genetics:
 A/Introns: 29/1; 95/3; 112/3; 150/3; 175/3; 193/3; 211/3; 229/3; 247/3; 265/3; 283/3; 295/3; 673/3; 706/3; 742/3; 760/3; 776/3; 796/3; 814/3; 850/3; 868/3; 886/3; 940/3; 976/3
 C/Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
 C/Keywords: coiled coil; extracellular matrix
 F/1-24/Domain: signal sequence #status predicted <SIG>
 F/25-154/Domain: propeptide #status predicted <PRO>
 F/32-97/Domain: von Willebrand factor type C repeat homology <VWC>
 F/155-1464/Product: collagen alpha 1(I) chain #status predicted <MAT>
 F/1236-1464/Domain: fibrillar collagen carboxyl-terminal homology <FCC>

Query Match

15.2%; Score 293; DB 2; Length 1464;
 Best Local Similarity 29.5%; Pred. No. 5.1e-11;

Matches 113; Conservative 23; Mismatches 117; Indels 130; Gaps 18;

```

QY 1 GGRGMSGAGEGFGFKPGPMDEGPDLDGPPVDEDESDNSAIYVQGLNDSVTLDDL 60
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 437 GAKGEPGARERGEREAGSPGIPGPKGEDGKSGSPGEPANGLPGAGRGSPSG----- 488
QY 61 DFKGQGVVMMNRKTGQPMHLYLDEKTKGPKDQATVSDPPTAKAAVWFPDQFGS 120
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 489 --FR-----GPAAGNGIPGKGPGE-----507
QY 121 KLVSLAKRKPMMNSMRG--GLPRRGRGMPPLRGGGPGGPG--GPMGRMG--170
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 508 -----RGRPGAGPRGPGVAGFGRDTPGPGGIRGMGPGSGPGRNDGKRGPGSGSGS 560
QY 171 ----GRGDRG-----GPPRRGRGRNPGSGGVNVRHAGMOCNPGCGNQNPAMRT 220
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 561 RPPRPGSPGPGQPGVWGF--GPKGNDGAPGKNG--ERGGGPGGPGPLPGPAGKN----- 610
QY 221 EGNQCKAPKPEGLPFPFP--PGDNRGPGGPGMG--GRG-----259
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 611 ----GETPQG--PGPGRPADKDSPPPGQIGTGTGPPGGRGKGEPPGPK 662
QY 260 -LMDRGPGMFRGGRGDRG-----GFRGGRGMDRGFGGRRGPGGPPG--P 306
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 663 EVGAPGAPGKGDSPGAGRGPRGTAGTPGARGGAG--PPRPGGKGPAGPPPGPAGSGSP 721
QY 307 LMEQMGRRGGRGPG--KMDKG 328
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 722 GLQGMGERGPGSPGPKGKGE 744

```

RESULT 14

T33925
 hypothetical protein Y66H1A.4 - Caenorhabditis elegans
 C/Species: Caenorhabditis elegans
 C/Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 09-Jul-2004
 A/Accession: T33925
 A/Reference number: Z21440
 A/Status: preliminary; translated from GB/EMBL/DBJ
 A/Molecule type: DNA
 A/Residues: 1-276 <CIA>
 A/Cross-references: UNIPROT:Q9TYK1; EMBL:AF125462; PIDN:AA12858.1; GSPDB:GN00022; CESP:
 C/Genetics:
 A/Experimental source: strain Bristol N2, clone Y66H1A
 A/Gene: CESP:Y66H1A.4

A/Map position: 4
 A/Introns: 12/3; 143/3
 C/Superfamily: Phaseolus glycine-rich cell wall protein 1.8

Query Match 15.1%; Score 291; DB 2; Length 276;
 Best Local Similarity 38.7%; Pred. No. 1.5e-11;

Matches 92; Conservative 6; Mismatches 60; Indels 80; Gaps 11;

```

QY 143 REGGRMPPLRGGRPGGPGPMGRMGRRGDRGFPFRGGRGSGNPNVHRAG 202
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 36 RGRGGRGGGFRGGRGGGG--GFRGGRGDRG--GFRGGRGGGGGGYDQG 87
QY 203 -----DMQCPNP-----GCNQNPAMR 219
DB 88 PPEBVVUVFVSHOCDDIVCNNTSGKLYFNAPYFKKQVGTIDIFGSGENGFSV 147
QY 220 TEQNOCKA--PKRGLPFP-----FPPGDRG--GPGMRGGRGLMDRG 265
DB 148 TISQGVKASFEBSGLYIDPKLLPVDRFLPQAGGGRGGRGDRGSGS--DRG 205
QY 266 PGGMFRG--RGDRGFRGGRGMDRGFGGRRGPGPPLMEQMGRRGGRG 319
DB 206 RGRGGRGGGGRGDRGGRGGRGGRGGRGGRGGRGGRGGRGGRG--DFGR--GRG 254

```

RESULT 15

CCGH15
 collagen alpha 1(I) chain - chicken (tentative sequence) (fragmente)

C/Species: Gallus gallus (chicken)

C/Date: 12-Aug-1981 #sequence_revision 06-Jul-1982 #text_change 31-Mar-2000

C/Accession: A90458; A90181; A02857

R/Reference: J.H.; Corbett, C.; Dixit, S.N.; Yu, W.; Seyer, J.M.; Kang, A.H.; Grose, J

Biochemistry 21, 2048-2055, 1982

A/Title: Amino acid sequence of chick skin collagen alpha1(I)-C88 and the complete prim

A/Reference number: A90458; MUID:82231995; PMID:7093229

A/Accession: A90458

A/Molecule type: protein

A/Residues: 1-1036 <HC>

A/Experimental source: skin

A/Note: this is the latest in a series of papers from these workers elucidating the seq

R/By: D.R.; Glimcher, M.J.

Biochem. Biophys. Res. Commun. 48, 720-726, 1972

A/Title: Evidence for a previously undetected sequence at the carboxyterminus of the al

A/Reference number: A90181; MUID:72243016; PMID:5047697

A/Accession: A90181

A/Molecule type: protein

A/Residues: 1037-1042 <EY>

A/Experimental source: skin

A/Note: residues 1037-1042 above correspond to the carboxyl end of the protein

C/Comment: Lysines at positions 103, 700, 934, and 946 above may be hydroxylated in som

C/Comment: Most of the prolines at the third position of the tripeptide repeating unit

C/Comment: Pro-1002 is the only 3-hydroxyproline and the only hydroxylated proline in p

C/Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology

C/Keywords: coiled coil; extracellular matrix; glycoprotein; pyroglyutamic acid; trimer;

F/1/Modified site: pyrrolidone carboxylic acid (Gln) #status experimental

Query Match 14.8%; Score 283.5; DB 1; Length 1042;
 Best Local Similarity 29.6%; Pred. No. 1.5e-10;

Matches 110; Conservative 17; Mismatches 118; Indels 127; Gaps 18;

```

QY 4 GGMGSGERGFENKPGPMDEGPDLDGPPVDEDESDNSAIYVQGLNDSVTLDDL 63
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 17 GPMGRGPRGLPFPAP--GPGGQGFPPGEGGAGSPMRG-----59
QY 64 KQGVVMMNRKTGQPMHLYLDEKTKGPKDQATVSDPPTAKAAVWFPDQFGSK 123
DB 60 -PAGPRGKGGDQ-----EAGKP--GRPGRGPPGGA-----RGLPGT 96
QY 124 VSLAKRKPMMNSMRG--GLPRRGRGMPPLRGGGPGGPGGPMGRMG--GDRGR 179
DB 97 AGI---PMKGRGRGSGLDGAKGPGPAPKPEPSPEAGAP--QMGRRGPGGERGP 151
QY 180 PPRGPRGSGNPGSGGVNVRHAGDMQCPNPGCGNQNPARTGNCCKAPKPEGLPFP 239

```

us-10-791-017a-2_copy_319_656.rpr

Page 9

Page 9

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using SW model

Run on: February 18, 2005, 15:13:56 ; Search time 170 Seconds

(without alignments)
1018.134 Million cell updates/sec

Title: US-10-791-017a-2_COPY_319_656
Perfect score: 1922
Sequence: 1 GGGGCGMSAGRGCFKPGG.....GGPGKDKGHRGRRDRY 338

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : UniProt_03:*
1: uniprot_sprot:*
2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1932	100.0	600	2 096MN4	096MN4 homo sapien
2	1932	100.0	600	2 096MN4	096MN4 homo sapien
3	1907.5	99.2	655	2 096TE8	096TE8 homo sapien
4	1907.5	99.2	661	2 096MX4	096MX4 homo sapien
5	1903	99.0	656	2 06NVA3	06NVA3 mus musculu
6	1888.5	98.3	655	1 EMS MOUSE	061545 mus musculu
7	1888.5	98.3	655	2 09D3P0	09D3P0 mus musculu
8	1874	97.5	655	2 09CRS5	09CRS5 mus musculu
9	1874	97.5	655	2 09CRS5	09CRS5 mus musculu
10	1874	97.5	655	2 09CRS5	09CRS5 mus musculu
11	1874	97.5	655	2 09CRS5	09CRS5 mus musculu
12	1874	97.5	655	2 09CRS5	09CRS5 mus musculu
13	1874	97.5	655	2 09CRS5	09CRS5 mus musculu
14	1874	97.5	655	2 09CRS5	09CRS5 mus musculu
15	1874	97.5	655	2 09CRS5	09CRS5 mus musculu
16	1874	97.5	655	2 09CRS5	09CRS5 mus musculu
17	1874	97.5	655	2 09CRS5	09CRS5 mus musculu
18	1874	97.5	655	2 09CRS5	09CRS5 mus musculu
19	1874	97.5	655	2 09CRS5	09CRS5 mus musculu
20	1874	97.5	655	2 09CRS5	09CRS5 mus musculu
21	1874	97.5	655	2 09CRS5	09CRS5 mus musculu
22	1874	97.5	655	2 09CRS5	09CRS5 mus musculu
23	1874	97.5	655	2 09CRS5	09CRS5 mus musculu
24	1874	97.5	655	2 09CRS5	09CRS5 mus musculu
25	1874	97.5	655	2 09CRS5	09CRS5 mus musculu
26	1874	97.5	655	2 09CRS5	09CRS5 mus musculu
27	1874	97.5	655	2 09CRS5	09CRS5 mus musculu
28	1874	97.5	655	2 09CRS5	09CRS5 mus musculu
29	1874	97.5	655	2 09CRS5	09CRS5 mus musculu
30	1874	97.5	655	2 09CRS5	09CRS5 mus musculu
31	1874	97.5	655	2 09CRS5	09CRS5 mus musculu

32	584.5	30.4	442	2 07PUC2	07PUC2 anopheles g
33	450.5	23.4	448	2 018265	018265 caenorhabdi
34	427	22.2	410	2 07ZUE3	07ZUE3 brachydanio
35	423	22.0	414	2 069TN3	069TN3 oryza sativ
36	414	21.5	394	2 06Y882	06Y882 oryza sativ
37	407.5	21.2	280	2 091VQ2	091VQ2 mus musculu
38	391	20.3	372	2 09AST1	09AST1 mus musculu
39	372	19.4	360	2 0691Y8	0691Y8 arabidopsis
40	370.5	19.3	211	2 09S5X7	09S5X7 arabidopsis
41	351	18.3	220	2 07PKU2	07PKU2 anopheles g
42	351	18.3	535	2 0942Q2	0942Q2 oryza sativ
43	338	17.6	157	2 070T18	070T18 homo sapien
44	311.5	16.2	423	2 09FM47	09FM47 homo sapien
45	311.5	16.2	604	2 091252	091252 streptomyce

ALIGNMENTS

RESULT 1	PRELIMINARY;	PRT;	600 AA.
096MN4	096MN4		
AC	01-DEC-2001 (T-EMBLrel. 19, Created)		
DT	01-DEC-2001 (T-EMBLrel. 19, Last sequence update)		
DT	01-MAR-2004 (T-EMBLrel. 26, Last annotation update)		
DE	Hypothetical protein FLJ32119.		
OS	Homo sapiens (human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
OX	NCBI_TaxID=9606;		
RA	[1]		
RP	SEQUENCE FROM N.A.		
RX	PDBmed=14702039; DOI=10.1038/ng1285;		
RA	Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,		
RA	Wakamatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,		
RA	Sekine M., Ohtsuka M., Nishi T., Shibahara T., Tanaka T., Ishii S.,		
RA	Yamamoto Y., Saito K., Kawai Y., Isono Y., Nakamura Y., Nagahara K.,		
RA	Murakami K., Yasuda T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,		
RA	Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa E., Omura Y.,		
RA	Abe K., Kamihara K., Katagata N., Sato K., Tanikawa M., Yamazaki M.,		
RA	Nishimura K., Ichihasi T., Yamashita H., Murakawa K., Fujimori K.,		
RA	Tanai H., Kimura M., Watanabe S., Hiroda S., Chiba Y., Iehida S.,		
RA	Ono Y., Takiguchi S., Watanabe S., Yoshida M., Hotta T., Kusano J.,		
RA	Kanehori K., Takahashi-Fujii A., Hara H., Tanase T., Nomura Y.,		
RA	Togiyasu S., Komai F., Hara R., Takeuchi K., Arita M., Imose N.,		
RA	Musashino K., Yuki H., Oshima A., Sasaki N., Aotsuka S.,		
RA	Yoshikawa Y., Matsunawa H., Ichihara T., Shichida N., Sano S.,		
RA	Moriya S., Momiyama H., Satoh N., Takami S., Terashima Y., Suzuki O.,		
RA	Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H.,		
RA	Hishigaki H., Watanabe T., Sugiyama A., Takemoto M., Kawakami B.,		
RA	Yamazaki M., Watanabe K., Kumagai A., Itakura S., Fukunaga Y.,		
RA	Fujimori Y., Komiyama M., Tashiro H., Taniguchi A., Fujiwara T.,		
RA	Ono T., Yamada K., Fujii Y., Ozaki K., Hiro M., Omori Y.,		
RA	Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,		
RA	Okitani R., Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T.,		
RA	Matsunuma K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M.,		
RA	Togashi T., Oyama M., Hata H., Watanabe M., Komatsu T.,		
RA	Mizushima-Sugano J., Satoh T., Shirai Y., Takahashi Y., Nakagawa K.,		
RA	Okumura K., Nagase T., Nomura N., Kikuchi H., Masuno Y., Yamashita R.,		
RA	Nakai K., Yada T., Nakamura Y., Ohara O., Isegaki T., Sugano S.,		
RT	"Complete sequencing and characterization of 21,243 full-length human		
RT	cDNAs."		
RL	Nat. Genet. 36:40-45(2004).		
EMBL	AK056681; BAB71252.1; -		
HSSP	095218; 1N0Z.		
DR	GO:0005634; C:nucleus; IEA.		
DR	GO:0005676; F:nucleic acid binding; IEA.		
DR	GO:0008270; P:zinc ion binding; IEA.		
DR	InterPro: IPR011368; RNA-binding_EWS.		
DR	InterPro: IPR000504; RNA_rec_mot.		
DR	InterPro: IPR01876; znf_RanGDP.		

DR Pfam: PF00076; RRM 1; 1.
 DR Pfam: PF00641; zf-RanBP; 1.
 DR PIRSF: PIRSF002101; RNA-binding_EWS; 1.
 DR SMART: SM00360; RRM; 1.
 DR SMART: SM00547; znf_RBZ; 1.
 DR PROSITE: PS50102; RRM; 1.
 DR PROSITE: PS01358; zf_RANBP2_1; 1.
 DR PROSITE: PS50199; zf_RANBP2_2; 1.
 SO SEQUENCE 600 AA; 62478 MW; 15720CBFC6023912 CRC64;

Query Match 100.0%; Score 1922; DB 2; Length 600;
 Best local similarity 100.0%; Pred. No. 3,6e-103;
 Matches 338; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGGGMSAGRGGFFKRGGPMDEGPDLDGPPDPDESDNATVYQGLNDSVTLLDIA 60
 263 GGGGMSAGRGGFFKRGGPMDEGPDLDGPPDPDESDNATVYQGLNDSVTLLDIA 322
 DB 61 DFRKCGVVMKRTGQPMHIYLDKGTGKGDATVSYEDPTAKAAVEMFGKDPQGS 120
 323 DFRKCGVVMKRTGQPMHIYLDKGTGKGDATVSYEDPTAKAAVEMFGKDPQGS 382
 QY 121 KLVSLARKKPPMNSMRGGLPREGRGMPPLRGPGGPGGPGMGRMGGRGDRGFP 180
 383 KLVSLARKKPPMNSMRGGLPREGRGMPPLRGPGGPGGPGMGRMGGRGDRGFP 442
 QY 181 PRPRGRGRNPGSGGNVQHRAGDMQCNPRGCGNPFAMRTGCTQCAPKREGLPPFP 240
 443 PRPRGRGRNPGSGGNVQHRAGDMQCNPRGCGNPFAMRTGCTQCAPKREGLPPFP 502
 DB 241 PGDGRGGRGPGMGRGGLMDRGPGGPMGRGGRGDRGPGGPGGGRGGRG 300
 503 PGDGRGGRGPGMGRGGLMDRGPGGPMGRGGRGDRGPGGPGGGRGGRG 562
 QY 301 GGGPPLMEQMGRRGGRGPGGRMDKGEHRQERRDRPY 338
 563 GGGPPLMEQMGRRGGRGPGGRMDKGEHRQERRDRPY 600
 DB

RESULT 2
 EWS_HUMAN STANDARD; PRT; 656 AA.
 ID 1 EWS_HUMAN
 AC 001844; Q92635;
 DT 01-JUN-1994 (Ref. 29, Created)
 DT 01-JUN-1994 (Ref. 29, Last sequence update)
 DT 23-OCT-2004 (Ref. 45, Last annotation update)
 DE RNA-binding protein EWS (EWS oncogene) (Ewing sarcoma breakpoint region 1 protein).
 DE Name=EWSR1; Synonyms=EWS;
 OS Homo sapiens (human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 OC NCBI_TaxID=9606;
 OX 1
 RN 1
 RP TISSUE=Brain; Fetal brain;
 RC MEDLINE=92396233; PubMed=1522903; DOI=10.1038/359162a0;
 RA Delattre O., Zucman J., Ploegastel B., Demaze C., Melot T., Peter M., Kovar H., Joubert I., de Jong P., Rouleau G., Aurias A., Thomas G.;
 RT "Gene fusion with an ETS DNA-binding domain caused by chromosome translocation in human tumours.";
 RT Nature 359:162-165(1992).
 RL 12
 RN 12
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94140360; PubMed=8307570;
 RA Ploegastel B., Zucman J., Peter M., Thomas G., Delattre O.;
 RT "Genomic structure of the EWS gene and its relationship to EWSR1, a site of tumor-associated chromosome translocation.";
 RT Genomics 18:609-615(1993).
 RL 13
 RN 13
 RP SEQUENCE FROM N.A.
 RA Zucman-Rossi J.;
 RT Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.

RN 14
 RP TISSUE=Placenta;
 RC MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
 RX Klausner R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shemmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Bucof K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heish F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Donald M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.U., Ueda T.B., Toshiyuki S., Carninci P., Franke C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abremson R.D., Mullany S.J.,
 RA Bosak S.A., McMan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Huylk S.W.,
 RA Villalon D.K., Muzny K.C., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahy J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whitting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butlerfield Y.S.N., Krzywinski M.J., Skalska U., Smolus D.E.,
 RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN 15
 RP SEQUENCE OF 1-345 FROM N.A.
 RX MEDLINE=97131501; PubMed=8975699; DOI=10.1006/geno.1996.0625;
 RA Zucman-Rossi J., Legoux P., Thomas G.;
 RT "Identification of new members of the Gae2 and Ras families in the 22q12 chromosome region.";
 RL Genomics 38:247-254(1996).
 RN 16
 RP SEQUENCE OF 241-268 FROM N.A.
 RC TISSUE=Placenta;
 RX MEDLINE=95352541; PubMed=7542907;
 RA Bhagrat T., Abe S., Nojima T., Yoshida M.C.;
 RT "Molecular analysis of a t(11;22) translocation junction in a case of Ewing's sarcoma.";
 RL Genes Chromosomes Cancer 13:126-132(1995).
 RN 17
 RP PHOSPHORYLATION SITE SER-266, AND PARTIAL SEQUENCE.
 RX MEDLINE=98001723; PubMed=9341188; DOI=10.1074/jbc.272.43.27369;
 RA Deloume J.C., Pritchard L., Delattre O., Storm D.R.;
 RT "The pronucleon protein EWS binds calmodulin and is phosphorylated by protein kinase C through an IQ domain.";
 RL J. Biol. Chem. 272:27369-27377(1997).
 RN 18
 RP SEQUENCE OF 128-158; 233-247; 268-324; 334-364; 393-439; 447-518 AND 551-641, METHYLATION OF ARGININES, AND MASS SPECTROMETRY.
 RX MEDLINE=21276345; PubMed=11278906; DOI=10.1074/jbc.M011446200;
 RA Belyanskaya L.L., Gehrig P.M., Gehrig H.;
 RT "Exposure on cell surface and extensive arginine methylation of Ewing sarcoma (EWS) protein.";
 RL J. Biol. Chem. 276:18681-18687(2001).
 RN 19
 RP ALTERNATIVE SPLICING, AND RNA-BINDING.
 RX MEDLINE=9436763; PubMed=8084618;
 RA Ohno T., Ouchida M., Lee L., Catalica Z., Rao V.N., Reddy E.S.P.;
 RT "The EWS gene, involved in Ewing family of tumors, malignant melanoma of soft parts and desmoplastic small round cell tumors, codes for an RNA binding protein with novel regulatory domains.";
 RL Oncogene 9:3087-3097(1994).
 RN 110
 RP INTERACTION WITH SPL.
 RX MEDLINE=98325009; PubMed=9660765; DOI=10.1074/jbc.273.29.18086;
 RA Zhang D., Paley A.J., Childs G.;
 RT "The transcriptional repressor ZFM1 interacts with and modulates the ability of EWS to activate transcription.";
 RL J. Biol. Chem. 273:18086-18091(1998).
 RN 111
 RP CHARACTERIZATION.
 RX MEDLINE=20390060; PubMed=10767297; DOI=10.1074/jbc.M002961200;
 LA K.K.C., Lee K.A.W.;

"Transcriptional activation by the Ewing's sarcoma (EWS) oncogene can be cis-repressed by the EWS RNA-binding domain.",
 RT J. Biol. Chem. 275:23053-23058(2000).
 CC -1- FUNCTION: Might normally function as a repressor. EWS-fusion-proteins (EFPs) may play a role in the tumorigenic process. They may disturb gene expression by mimicking, or interfering with the normal function of CTD-Polr1 within the transcription initiation complex. They may also contribute to an aberrant activation of the fusion protein target genes.
 CC -1- SUBUNIT: Binds POLR2C, SFI, calmodulin and RNA. Interacts with PRK2B/PKX2.
 CC -1- SUBCELLULAR LOCATION: Nuclear, cytoplasmic and on cell surface. Relocates from cytoplasm to ribosomes upon PRK2/PKX2 activation.
 CC -1- ALTERNATIVE PRODUCTS: Named isoforms=2;
 CC Name=EWS;
 CC IsoId=Q01844-1; Sequence=Displayed;
 CC Name=EWS-B;
 CC IsoId=Q01844-2; Sequence=VSP_005793;
 CC Note=No experimental confirmation available;
 CC -1- TISSUE SPECIFICITY: Ubiquitous.
 CC -1- DOMAIN: EWS activation domain (EAD) functions as a potent activation domain in EFPs. EWS1 binds POLR2C but not POLR2B or POLR2G, whereas the isolated EAD binds POLR2E and POLR2G but not POLR2C. Cis-linked RNA-binding domain (RBD) can strongly and specifically repress trans-activation by the EAD.
 CC -1- PTM: Phosphorylated; calmodulin-binding inhibits phosphorylation of Ser-266.
 CC -1- DISORDER: Ewing's sarcoma is characterized by chromosomal translocations t(11;22)(q24;q12) which involves EWSR1 and FLI1, t(7;22)(p22;q12) which involves EWSR1 and ETV1, t(21;22)(q22;q12) which involves EWSR1 and ERG and t(9;22)(q22-31;q11-12) which involves EWSR1 and NR4A3.
 CC -1- DISORDER: Involved in desmoplastic small round cell tumor (DSRCT) through a chromosomal translocation t(11;22)(p13;q12) that involves EWSR1 and WT1.
 CC -1- DISORDER: Malignant melanoma of soft parts (MMSP), also known as soft tissue clear cell sarcoma, is a rare tumor developing in tendons and aponeuroses. It is associated with chromosomal translocation t(12;22)(q13;q12) involving EWSR1 and ATF-1.
 CC -1- DISORDER: Involved in small round cell sarcoma through a chromosomal translocation t(1;22)(p36.1;q12) that involves EWSR1 and ZNF279.
 CC -1- MISCELLANEOUS: EFPs arise due to chromosomal translocations in which EWSR1 is fused to a variety of cellular transcription factors. EFPs are very potent transcriptional activators dependent on the EAD and a C-terminal DNA-binding domain contributed by the fusion partner. The spectrum of malignancies associated with EFPs are thought to arise via EFP-induced transcriptional deregulation, with the tumor phenotype specified by the EWSR1 fusion partner and cell type. Transcriptional repression of the transforming growth factor beta type II receptor (TGF beta RII) is an important target of the EWS-FLI1, EWS-ERG, or EWS-ETV1 oncogene.
 CC -1- MISCELLANEOUS: Binds calmodulin in the presence, but not in the absence, of calcium ion.
 CC -1- SIMILARITY: Belongs to the RNP TET family.
 CC -1- SIMILARITY: Contains 1 IQ domain.
 CC -1- SIMILARITY: Contains 1 RanBP2-type zinc finger.
 CC -1- SIMILARITY: Contains 1 RNA recognition motif (RNM) domain.
 CC -1- DATABASE: NAME=Atlas Genet. Cytogenet. Oncol. Haematol.; WWW="http://www.infobiogen.fr/services/chromancer/Genes/EWSR11085.html".
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
 CC EMBL; X66899; CAA47350.1; -
 CC EMBL; X72990; CAA51489.1; -
 CC EMBL; X72991; CAA51489.1; JOINED.

DR EMBL; X72992; CAA51489.1; JOINED.
 DR EMBL; X72993; CAA51489.1; JOINED.
 DR EMBL; X72994; CAA51489.1; JOINED.
 DR EMBL; X72995; CAA51489.1; JOINED.
 DR EMBL; X72996; CAA51489.1; JOINED.
 DR EMBL; X72997; CAA51489.1; JOINED.
 DR EMBL; X72998; CAA51489.1; JOINED.
 DR EMBL; X72999; CAA51489.1; JOINED.
 DR EMBL; X73000; CAA51489.1; JOINED.
 DR EMBL; X73001; CAA51489.1; JOINED.
 DR EMBL; X73002; CAA51489.1; JOINED.
 DR EMBL; X73003; CAA51489.1; JOINED.
 DR EMBL; X73004; CAA51489.1; JOINED.
 DR EMBL; BC004817; AA04817.1; -
 DR EMBL; Y08806; CAA70044.1; ALT_INIT.
 DR EMBL; AB016435; BAA31990.1; -
 DR EMBL; Y07848; CAA69177.1; -
 DR F1R; A49358; A49358.
 DR HSSP; 095218; 1N0Z.
 DR GeneW; HGNC.3508; EWSR1.
 DR H-INVD; HIX0016349; -
 DR MIM; 133450; -
 Query Match 100.0%; Score 1922; DB 1; Length 656;
 Best Local Similarity 100.0%; Pred. No. 3.9e-103;
 Matches 338; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 GGRGWSAGRGPGKPGPMDGPDLDLGPVDPDESDNSAIVVGLNSVTLLDIA 60
 DB 319 GGRGWSAGRGPGKPGPMDGPDLDLGPVDPDESDNSAIVVGLNSVTLLDIA 378
 QY 61 DFFKQGVVVRNKRITQPMIHYLYDKETGPKKDATVSIJEDPTAKAAVWPFQKPOGS 120
 DB 379 DFFKQGVVVRNKRITQPMIHYLYDKETGPKKDATVSIJEDPTAKAAVWPFQKPOGS 438
 QY 121 KLVSLARKKKPPNNSRGGLPPEGRGMPPLRGSGPGSGPGMGRGGRGDRGGP 180
 DB 439 KLVSLARKKKPPNNSRGGLPPEGRGMPPLRGSGPGSGPGMGRGGRGDRGGP 498
 QY 181 PGRGSRGNPGSGGVVQHRAGDMQCPNPGCGNPFMTETCNOCKAPKPEGLPPFP 240
 DB 499 PGRGSRGNPGSGGVVQHRAGDMQCPNPGCGNPFMTETCNOCKAPKPEGLPPFP 558
 QY 241 PGDRGRGPGGMRGGRGGLMDRGPGGMFRGGRGDRGFRGGRGDRGFRGGRG 300
 DB 559 PGDRGRGPGGMRGGRGGLMDRGPGGMFRGGRGDRGFRGGRGDRGFRGGRG 618
 QY 301 GGPGRPLMEOMGRRGGRGGRGPGKMGKGEHRRDRPPY 338
 DB 619 GGPGRPLMEOMGRRGGRGGRGPGKMGKGEHRRDRPPY 656
 RESULT 3
 Q96FE8 PRELIMINARY; PRT; 655 AA.
 AC Q96FE8;
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DT 25-OCT-2004 (TREMBLrel. 28, Last annotation update)
 DE Ewing sarcoma breakpoint region 1, isoform EWS (EWSR1 protein).
 GN Name=EWSR1;
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_Taxid=9606;
 RX [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Lymph, and Skin;
 RX MEDLINE=2388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klaunder R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,

[illegible]

Db 558 PGDBRGRCGPGMGRGRCGLMDRGGPFRGRCGRGDRGFGRCGRMDGSGFCGCRRGCP
 Qy 301 GGPPEPLMEQNGRRGRCGRGRCRKNKCGEHRORRRRPY 338
 Db 618 GGPPEPLMEQNGRRGRCGRGRCGRKMDGGEHRORRRRPY 655
 RESULT 4
 Q96MX4 PRELIMINARY; PRT; 661 AA.
 AC Q96MX4;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 OS Hypothetical protein FLJ31747.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 NCBI_TaxId=9606;
 [1]
 SEQUENCE FROM N.A.
 RP PubMed=14702039; DOI=10.1038/ng1285;
 RX Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,
 RA Wakematsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,
 RA Sekine M., Ohyaishi M., Nishi T., Shibahara T., Tanaka T., Itoh S.,
 RA Yamamoto J., Saito K., Kawai Y., Isono Y., Nakamura Y., Nagahari K.,
 RA Murakami K., Yasuda T., Iwayanagi T., Wagatsuma M., Shiratori A.,
 RA Sudo H., Hosobri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
 RA Takehashi M., Kanda K., Yokoi T., Furuya T., Kikkawa E., Omura Y.,
 RA Abe K., Kamihira K., Katsuta N., Sato K., Tanikawa M., Yamazaki M.,
 RA Niimiyama K., Ishibashi T., Yamashita K., Murakawa K., Fujimori K.,
 RA Tanai H., Kimura M., Watanabe M., Hirakawa S., Chiba Y., Iehida S.,
 RA Ono Y., Takiguchi S., Watanabe S., Yosida M., Hotuta T., Kusano J.,
 RA Kanehori K., Takahashi-Fujii A., Hara H., Tanase T., Nomura Y.,
 RA Togaya S., Komai F., Hara R., Takeuchi K., Arita M., Imose N.,
 RA Yoshishino K., Yuzuki H., Oshima A., Sasaki N., Aotsuma S.,
 RA Yoshikawa Y., Matsunawa H., Ichihara T., Shiohara N., Sano S.,
 RA Moriya S., Momiyama H., Satoh N., Takami S., Terashima Y., Suzuki O.,
 RA Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakabe H.,
 RA Hishigaki S., Watanabe T., Sugiyama A., Takemoto M., Kawakami B.,
 RA Yamazaki M., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y.,
 RA Fujimori Y., Komiyama M., Tashiro H., Tanigami A., Fujiwara T.,
 RA Ono T., Yamada K., Fujii Y., Okaki K., Hirao M., Omori Y., Okamoto S.,
 RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
 RA Ohtani R., Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T.,
 RA Matsumura K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M.,
 RA Togoishi T., Oyama M., Hata H., Watanabe M., Komatsu T.,
 RA Mizushima-Sugano J., Satoh T., Shirai Y., Takahashi Y., Nakagawa K.,
 RA Okumura K., Nagae T., Nomura N., Kikuchi H., Masuno Y., Yamashita K.,
 RA Nakai K., Yada T., Nakamura Y., Ohara O., Isegaki T., Sugano S.,
 RT "Complete sequencing and characterization of 21,243 full-length human
 CDNA's".
 RT Nat. Genet. 36:40-45(2004).
 DR EMBL; AK056309; BAB71145.1; -.
 DR HSSP; O95218; INOZ.
 DR GO; GO:0005634; C:nucleus; IEA.
 DR GO; GO:0003676; F:nucleic acid binding; IEA.
 DR GO; GO:0008270; F:zinc ion binding; IEA.
 DR InterPro; IPR011368; RNA-binding_EWS.
 DR InterPro; IPR000504; RNA_rec_mot.
 DR InterPro; IPR001876; ZnF_RanGDP.
 DR Pfam; PF00076; RRM_1; 1.
 DR Pfam; PF00641; ZF_RanBP; 1.
 DR PIRSF; PIRSF002101; RNA-binding_EWS; 1.
 DR SMART; SM00360; RRM; 1.
 DR SMART; SM00547; ZNF_RBZ; 1.
 DR PROSITE; PS01012; RRM; 1.
 DR PROSITE; PS01358; ZF_RANBP2_1; 1.
 DR PROSITE; PS01019; ZF_RANBP2_2; 1.
 SQ SEQUENCE 661 AA; 68966 MW; 5F84F52FDDC051269 CRC64;

Best Local Similarity 99.7%; Pred. No. 2,7e-102; Matches 337; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 GGRGMSAGRGGFPKPGPMDEGPDLDGPVDPEDSDNSAIYVQGLNDSTVTLDDL 60
 DB 325 GGRGMSAGRGGFPKPGPMDEGPDLDGPVDPEDSDNSAIYVQGLNDSTVTLDDL 383
 QY 61 DFFKQGVVNMNRRTGQPMIHIYLDKETGPKGDATVSYEDPPTAKAAVWFDGKDFQGS 120
 DB 384 DFFKQGVVNMNRRTGQPMIHIYLDKETGPKGDATVSYEDPPTAKAAVWFDGKDFQGS 443
 QY 121 KLVSLARKKPPNNSNRGGLPPREGMPPLRLGPGGPGGPGMGRMGGRGDRGFP 180
 DB 444 KLVSLARKKPPNNSNRGGLPPREGMPPLRLGPGGPGGPGMGRMGGRGDRGFP 503
 QY 181 PRGPRGSRGNPSGGGNVQHRAGDMQCPNPGCQGNFAMRTECNOCKAPKEGFLPPFP 240
 DB 504 PRGPRGSRGNPSGGGNVQHRAGDMQCPNPGCQGNFAMRTECNOCKAPKEGFLPPFP 563
 QY 241 PGDGRGRGPGGMRGGRGGLMDRGPGGMRGGRGGRGGRGGRGGRGGRGGRG 300
 DB 564 PGDGRGRGPGGMRGGRGGLMDRGPGGMRGGRGGRGGRGGRGGRGGRGGRG 623
 QY 301 GGPPLMEQMGRRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRG 338
 DB 624 GGPPLMEQMGRRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRG 661

RESULT 5
 ID 06NVA3 PRELIMINARY; PRT; 656 AA.

AC 06NVA3;
 DT 05-JUL-2004 (TEMBLrel. 27, Created)
 DT 05-JUL-2004 (TEMBLrel. 27, Last sequence update)
 DE Ewing sarcoma breakpoint region 1.
 GN Name=Ewari;
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6; TISSUE=Brain;
 RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klauener R.D., Collins P.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buettow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh P.,
 RA Diatchenko U., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Ueidi T.B., Tothilyuk I., Camarillo P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
 RA Krzywicki M.I., Skalka U., Smalins D.E., Schmechel A., Schein J.B.,
 RA Jones S.J., Watra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 and mouse cDNA sequences";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6; TISSUE=Brain;
 RA Director MGC Project;
 RA Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC068226; AA68226.1; -;
 DR GO; GO:0005634; C:nucleus; IDA.
 DR InterPro; IPR011368; RNA-binding_EWS.

DR InterPro; IPR000504; RNA_rec_mot.
 DR InterPro; IPR001876; Znf_RanBDP.
 DR Pfam; PF00076; RRM_1; 1.
 DR Pfam; PF00641; zf_RanBP; 1.
 DR PIRSF; PIRSF002101; RNA-binding_Ews; 1.
 DR SMART; SM00360; RRM; 1.
 DR SMART; SM00547; Znf_RBZ; 1.
 DR PROSITE; PS0102; RRM; 1.
 DR PROSITE; PS01358; zf_RanBP2_1; 1.
 DR PROSITE; PS01399; zf_RanBP2_2; 1.
 SQ SEQUENCE 656 AA; 68549 MW; 5DF018A22C8C1D0B CRC64;

Query Match 99.0%; Score 1903; DB 2; Length 656;
 Best Local Similarity 98.5%; Pred. No. 4.9e-102;
 Matches 337; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGRGMSAGRGGFPKPGPMDEGPDLDGPVDPEDSDNSAIYVQGLNDSTVTLDDL 60
 DB 319 GGRGMSAGRGGFPKPGPMDEGPDLDGPVDPEDSDNSAIYVQGLNDSTVTLDDL 378
 QY 61 DFFKQGVVNMNRRTGQPMIHIYLDKETGPKGDATVSYEDPPTAKAAVWFDGKDFQGS 120
 DB 379 DFFKQGVVNMNRRTGQPMIHIYLDKETGPKGDATVSYEDPPTAKAAVWFDGKDFQGS 438
 QY 121 KLVSLARKKPPNNSNRGGLPPREGMPPLRLGPGGPGGPGMGRMGGRGDRGFP 180
 DB 439 KLVSLARKKPPNNSNRGGLPPREGMPPLRLGPGGPGGPGMGRMGGRGDRGFP 498
 QY 181 PRGPRGSRGNPSGGGNVQHRAGDMQCPNPGCQGNFAMRTECNOCKAPKEGFLPPFP 240
 DB 499 PRGPRGSRGNPSGGGNVQHRAGDMQCPNPGCQGNFAMRTECNOCKAPKEGFLPPFP 558
 QY 241 PGDGRGRGPGGMRGGRGGLMDRGPGGMRGGRGGRGGRGGRGGRGGRGGRG 300
 DB 559 PGDGRGRGPGGMRGGRGGLMDRGPGGMRGGRGGRGGRGGRGGRGGRGGRG 618
 QY 301 GGPPLMEQMGRRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRG 338
 DB 619 GGPPLMEQMGRRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRG 656

RESULT 6
 ID EWS_MOUSE STANDARD; PRT; 655 AA.

AC 061545;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 25-OCT-2004 (Rel. 45, Last annotation update)
 DE RNA-binding protein EWS.
 GN Name=Ewari; Synonym=EWS; Ewsh;
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Testis;
 RX MEDLINE=95130099; PubMed=7829090;
 RA Pionasael B., Mattei M.-G., Thomas G., Delattre O.;
 RT "Cloning and chromosome localization of the mouse Ews gene";
 RL Genomics 23:278-281(1994).
 CC -1- FUNCTION: Might function as a repressor (By similarity).
 CC -1- SUBUNIT: Binds RNA, POLR2C, SFI and calmodulin. Interacts with
 CC PTK2B (By similarity).
 CC -1- SUBCELLULAR LOCATION: Nuclear, cytoplasmic and on cell surface.
 CC Relocates from cytoplasm to ribosomes upon PTK2B/FKX2 activation
 CC (By similarity).
 CC -1- PTM: Phosphorylated; calmodulin-binding inhibits phosphorylation
 CC of Ser-266 (By similarity).
 CC -1- MISCELLANEOUS: Binds calmodulin in the presence, but not in the
 CC absence, of calcium ion (By similarity).
 CC -1- SIMILARITY: Belongs to the RNP TET family.
 CC -1- SIMILARITY: Contains 1 IQ domain.

FT		MOD_RES	309	309	Asymmetric dimethylarginine (By similarity).
FT		MOD_RES	314	314	Asymmetric dimethylarginine (By similarity).
FT		MOD_RES	317	317	Asymmetric dimethylarginine (By similarity).
FT		MOD_RES	321	321	Asymmetric dimethylarginine (By similarity).
FT		MOD_RES	454	454	Asymmetric dimethylarginine (By similarity).
FT		MOD_RES	463	463	Asymmetric dimethylarginine (By similarity).
FT		MOD_RES	470	470	Asymmetric dimethylarginine (By similarity).
FT		MOD_RES	489	489	Asymmetric dimethylarginine (By similarity).
FT		MOD_RES	493	493	Asymmetric dimethylarginine (By similarity).
FT		MOD_RES	499	499	Asymmetric dimethylarginine (By similarity).
FT		MOD_RES	502	502	Asymmetric dimethylarginine (By similarity).
FT		MOD_RES	505	505	Asymmetric dimethylarginine (By similarity).
FT		MOD_RES	562	562	Asymmetric dimethylarginine (By similarity).
FT		MOD_RES	564	564	Asymmetric dimethylarginine (By similarity).
FT		MOD_RES	571	571	Asymmetric dimethylarginine (By similarity).
FT		MOD_RES	574	574	Asymmetric dimethylarginine (By similarity).
FT		MOD_RES	580	580	Asymmetric dimethylarginine (By similarity).
FT		MOD_RES	588	588	Asymmetric dimethylarginine (By similarity).
FT		MOD_RES	591	591	Asymmetric dimethylarginine (By similarity).
FT		MOD_RES	595	595	Asymmetric dimethylarginine (By similarity).
FT		MOD_RES	599	599	Asymmetric dimethylarginine (By similarity).
FT		MOD_RES	602	602	Asymmetric dimethylarginine (By similarity).
FT		MOD_RES	606	606	Asymmetric dimethylarginine (By similarity).
FT		MOD_RES	614	614	Asymmetric dimethylarginine (By similarity).
FT		MOD_RES	632	632	Asymmetric dimethylarginine (By similarity).
FT		MOD_RES	635	635	Asymmetric dimethylarginine (By similarity).
FT		MOD_RES	266	266	Phosphoserine (By PKC) (By similarity).
SEQ		SEQUENCE	655 AA;	66418 MW;	50735DEBDS4247D69 CRC64;
Query Match					98.3%; Score 1888.5; DB 1; Length 655;
Best Local Similarity					98.2%; Pred. No. 3.3e-101;
Matches 332; Conservative					4; Mismatches 1; Indels 1; Gaps 1;
QY	1	GGRGMSAGRGPNKPGPMDCGPLDLGPVDPDSDNSAIYYOGLNDSLTLDDLA	60		
DB	319	GGRGLG-AGRRGFNRPGGPMDCGPLDLGLPLDPEDSDNSAIYYOGLNDNVTLDDL	377		
QY	61	DFFPGCGVKKNKKTGGPMTIHYLDKETGKRKGATSYEDPPAKAAVEWFDGKD	120		
DB	378	DFFPGCGVKKNKKTGGPMTIHYLDKETGKRKGATSYEDPPAKAAVEWFDGKD	437		
QY	121	KLVKSLARKKPNNMSEGJLPREGRGWPEPLRGCPCGPGGPGPMGRMGSGDRGSFP	180		
DB	438	KLVKSLARKKPNNMSEGMPREGRGWPEPLRGCPCGPGGPGPMGRMGSGDRGSFP	497		
QY	181	PGRGRGSRGNPDGGGVNOHRAGMQCNPFGCGNQNFMARTECNCKAPKPEGFLPPPPP	240		

Db	498	PRGPRGSKGNBPSCGGNAGVQHRADDMQCPNCGQNTAMRTKCNQCTAPRREGTLPPEFP	557
Oy	241	PGDGRGCGPGMGRGCGGLMDRGSGGMFRGCGGDRGCGFRGCGMDRGCGGRRGCGP	3000
Db	558	PGDGRGCGPGMGRGCGGLMDRGSGGMFRGCGGDRGCGFRGCGMDRGCGGRRGCGP	617
Oy	301	GCGPGLMEQMGRRGCGRGCGRCGMDGCGHROERDRPY	338
Db	618	GCGPGLMEQMGRRGCGRGCGRCGMDGCGHROERDRPY	655
RESULT 7			
O9D250			
ID	O9D250	PRELIMINARY, PRT, 655 AA.	
AC	O9D250		
DT	01-JUN-2001	(TREMBLrel. 17, Created)	
DT	01-JUN-2001	(TREMBLrel. 17, last sequence update)	
DT	01-MAR-2004	(TREMBLrel. 26, last annotation update)	
DE	Mus musculus 18 days pregnant adult female placenta and extra embryonic tissue cDNA, RIKEN full-length enriched library, clone:3830417b11 product:Ewing sarcoma homolog, full insert sequence.		
DE	clone:3830417b11 product:Ewing sarcoma homolog, full insert sequence.		
GN	Name=Ewari; Synonyms=Eweh;		
OS	Mus musculus (Mouse);		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
OX	NCBI_TaxId=10090;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=C57BL/6J; TISSUE=Placenta and extra embryonic tissue;		
RX	MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;		
RA	Carninci P., Hayashizaki Y.;		
RT	"High-efficiency full-length cDNA cloning.";		
RL	Mech. Enzymol. 303:19-44(1999).		
RP	[2]		
RC	SEQUENCE FROM N.A.		
RX	STRAIN=C57BL/6J; TISSUE=Placenta and extra embryonic tissue;		
RA	MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;		
RT	RIKEN FANTOM Consortium;		
RL	"Functional annotation of a full-length mouse cDNA collection.";		
RL	Nature 409:685-690(2001).		
RP	[3]		
RC	SEQUENCE FROM N.A.		
RX	STRAIN=C57BL/6J; TISSUE=Placenta and extra embryonic tissue;		
RA	The FANTOM Consortium;		
RT	"The RIKEN Genome Exploration Research Group Phase I & II Team:		
RT	"Analysis of the mouse transcriptome based on functional annotation of		
RT	60,770 full-length cDNAs.";		
RL	Nature 420:563-573(2002).		
RP	[4]		
RC	SEQUENCE FROM N.A.		
RX	STRAIN=C57BL/6J; TISSUE=Placenta and extra embryonic tissue;		
RA	MEDLINE=20493974; PubMed=11042159; DOI=10.1101/gr.145100;		
RA	Komino H., Shihata Y., Hayatsu N., Sugahara Y., Shihata K., Itoh M.,		
RA	Carninci P., Hayashizaki Y.;		
RT	"Normalization and substructure of cap-trapper-selected cDNAs to		
RT	prepare full-length cDNA libraries for rapid discovery of new genes.";		
RL	Genome Res. 10:1617-1630(2000).		
RP	[5]		
RC	SEQUENCE FROM N.A.		
RX	STRAIN=C57BL/6J; TISSUE=Placenta and extra embryonic tissue;		
RX	MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;		
RA	Shihata K., Itoh M., Atawa K., Nagao S., Sasaki N., Carninci P.,		
RA	Sun H., Akiyama J., Nishi K., Katsunari T., Tashiro H., Itoh M.,		
RA	Sun H., Ishii Y., Nakamura S., Hazama M., Nishibe T., Hatada A.,		
RA	Fujimoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,		
RA	Fujimoto S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,		
RA	Ozawa Y., Ishikawa T., Ozawa K., Tanaka T., Matsura S., Kawai J.,		
RA	Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;		
RT	"RIKEN integrated sequence analysis (RISA) system-384-format		
RT	sequencing pipeline with 384 multicapillary sequencer.";		
RL	Genome Res. 10:1757-1771(2000).		

RN	[6]
RC	SEQUENCE FROM N.A.
RP	STRIN=C57BL/6J; TISSUE=Placenta and extra embryonic tissue;
RA	Adechi J., Atzawa K., Akahira S., Akimura T., Arai A., Aono H.,
RA	Arakawa T., Bono H., Carninci P., Fukunishi Y., Furuno M.,
RA	Hanagata T., Hara A., Hayatsu N., Hiramoto K., Hirooka T., Horii F.,
RA	Iinatsari K., Ishii Y., Itoh M., Itawa M., Kasukawa T., Kato H.,
RA	Kawai J., Kojima Y., Konno H., Kouda M., Koya S., Kurihara C.,
RA	Mateuyama T., Miyazaki A., Nishi K., Nomura K., Numazaki R., Ono M.,
RA	Okazaki Y., Okido T., Owa C., Saito H., Saito R., Sakai C., Sakai K.,
RA	Sano H., Saeki D., Shibata K., Shibata Y., Shingawa A., Shiraki T.,
RA	Sogabe Y., Suzuki H., Tagami M., Tagawa A., Takahashi F., Tanaka T.,
RA	Tejima Y., Toyota T., Yamamura T., Yasunishi A., Yoshida K., Yoshino M.,
RA	Mutamatsumu M., Hayashizaki Y.,
RL	Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
DR	EMBL; AK019460; BAB31732.1; -
DR	HSSP; O95218; INOZ.
DR	MGD; MG1:9960; Ewrl.
DR	GO; GO:0005634; C:nucleus, IDA.
DR	InterPro; IPRO11368; RNA-binding_EWS.
DR	InterPro; IPR005004; RNA_rec_mot.
DR	InterPro; IPR01876; Znf_RangBP.
DR	Pfam; PF00076; RM1_1; 1.
DR	Pfam; PF00641; zf-RanBP; 1.
DR	PIRSF; PIRSF002101; RNA-binding_EWS; 1.
DR	SMART; SM00360; RM1; 1.
DR	SMART; SM00547; Znf_RB2; 1.
DR	PROSITE; PS0102; RM; 1.
DR	PROSITE; PS01358; ZF_RANBP2_1; 1.
DR	PROSITE; PS50199; ZF_RANBP2_2; 1.
SQ	SEQUENCE 655 AA; 68462 MW; 10C7F06A92166873 CRC64;
Query Match	98.3%; Score 1888.5; DB 2; Length 655;
Best Local Similarity	98.2%; Pred.No.3.3e-101;
Matches 332; Conservative 4; Mismatches 1; Indels 1; Gaps 1;	
QY	1 GGRGMSAGSRGGGFNRKPGPMDEGPLDUGPYDPEDSDNSAIYYGLINDSVTLDDLA 60
DB	319 GGRGGLG-AGRGCGFNRRPGPMDEGPLDLGLPIDPDSDNSAIYYQLINDVNTLLDILA 377
QY	61 DFPKOCGVVKKRKTGTGQPMTHIYLDKETGPKGATASYDDPPAKAAVEPFQSKDQGS 120
DB	378 DFPKOCGVVKKRKTGTGQPMTHIYLDKETGPKGATASYDDPPAKAAVEPFQSKDQGS 437
QY	121 KLKYSLARKKPEPNMNMKGSLPREGRGMPLLRGCPGCGPGGPGMWGMGARGGDRGCFP 180
DB	438 KLKYSLARKKPEPNMNMKGMPRRGRGMPLLRGCPGCGPGGPGMWGMGARGGDRGCFP 497
QY	181 PRGRGSRGNDSGGSNVQHRAQWQCNPCCGNQNFAMRTENCNQCKAPKEGFLPPPPP 240
DB	498 PRGRGSRGNDSGGSNVQHRAQWQCNPCCGNQNFAMRTENCNQCKAPKEGFLPPPPP 557
QY	241 PGGRGRGCRPGMGNGRGGGLMDRGGPCGMRGGRGSGRGGFRGGRGNDRGGFGGRRGCGP 300
DB	558 PGGRGRGCRPGMGNGRGGGLMDRGGPCGMRGGRGSGRGGFRGGRGNDRGGFGGRRGCGP 617
QY	301 GGPPGPMLMEQWRGRRGGRGCGKMDKGEHQRERDRPY 338
DB	618 GGPPGPMLMEQWRGRRGGRGCGKMDKGEHQRERDRPY 655
RESULT 8	
Q9CRSS	PRELIMINARY; PRT; 333 AA.
AC	Q9CRSS; 01-JUN-2001 (TREMBlrel. 17, Created)
DT	01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT	01-MAR-2004 (TREMBlrel. 26, Last annotation update)
DE	Mus musculus 17 days embryo head cDNA, RIKEN full-length enriched library, clone:1300002011 product:Ewing sarcoma homolog, full insert sequence. (Fragment).
GN	Name=Ewrt1; Synonym=Ewbsh;
OS	Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OK NCBI_TaxID=10090;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Head;
 RX MEDLINE=9279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
 RA Carninci P., Hayashizaki Y.;
 RT "High-efficiency full-length cDNA cloning.";
 RL MGH. Enzymol. 303:19-44(1999).
 RN (2)
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Head;
 RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
 RA RIKEN PANTOM Consortium;
 RT "Functional annotation of a full-length mouse cDNA collection.";
 RL Nature 409:685-690(2001).
 RN (3)
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Head;
 RA The PANTOM Consortium;
 RT "Analysis of the mouse transcriptome based on functional annotation of
 60,770 full-length cDNAs.";
 RL Nature 420:563-573(2002).
 RN (4)
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Head;
 RX MEDLINE=20493374; PubMed=11042159; DOI=10.1101/gr.145100;
 RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
 RA Kono H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
 RT "Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new genes.";
 RL Genome Res. 10:1517-1530(2000).
 RN (5)
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Head;
 RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
 RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
 RA Kono H., Akiyama J., Nishi K., Kiteunai T., Taahiro H., Itoh M.,
 RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishino T., Harada A.,
 RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
 RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watanabe M.,
 RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Maruyama S., Kawai J.,
 RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
 RT "RIKEN integrated sequence analysis (RISA) system-384-format
 sequencing pipeline with 384 multicapillary sequencer.";
 RL Genome Res. 10:1757-1771(2000).
 RN (6)
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Head;
 RA Aichi J., Aizawa K., Akahira S., Akimura T., Arai A., Aono H.,
 RA Hanagaki T., Bono H., Hayatsu N., Hirakata S., Fukunishi Y., Furuno M.,
 RA Hanagaki T., Hara A., Hayatsu N., Hirakata S., Kuroki T., Hori F.,
 RA Imoto K., Ishii Y., Itoh M., Izawa M., Kasuwa T., Kato H.,
 RA Iwai Y., Kojima Y., Kono H., Kouda M., Koya S., Kuribara C.,
 RA Matsuyama T., Miyazaki A., Nishi K., Nomura K., Numazaki R., Ono M.,
 RA Okazaki Y., Okido T., Owa C., Saito H., Saito R., Sakai C., Sakai T.,
 RA Sano H., Sasaki D., Shibata K., Shibata Y., Shingawa A., Shiraki T.,
 RA Sogabe Y., Suzuki H., Tagami T., Tagawa A., Takahashi F., Tanaka T.,
 RA Tejima Y., Toya T., Yamamura T., Yasunishi A., Yoshida K., Yoshino M.,
 RA Muramatsu M., Hayashizaki Y.;
 RT Submitted (Jul-2000) to the EMBL/GenBank/DBJ databases.
 RL EMBL; AK014366; BAB29301.1; -.
 DR HSPF; 095218; INOZ.
 DR MGD; MGI:99960; Ewari.
 DR GO; GO:0005634; C:nucleus; IDA.
 DR InterPro; IPR000504; RNA_rec_mot.
 DR InterPro; IPR001875; znf_RandDP.
 DR Pfam; PF00076; RRM_1; 1.
 DR Pfam; PF00641; zf_RanBP; 1.
 DR SMART; SM00360; RRM; 1.
 DR SMART; SM00547; znf_RBZ; 1.

DR PROSITE; PS50102; RRM; 1.
 DR PROSITE; PS01358; zf_RANBP2_1; 1.
 DR PROSITE; PS50199; zf_RANBP2_2; 1.
 FT NON_TER 1
 FT TER 1
 SQ SEQUENCE 333 AA; 34617 MW; CAEDCA33F2F07 CRC64;
 Query Match 97.5%; Score 1874; DB 2; Length 333;
 Best Local Similarity 98.5%; Pred. No. 1.2e-100;
 Matches 328; Conservative 4; Mismatches 1; Indels 0; Gaps 0;
 QY 6 MSGAGRGFGNKGPGMDGPDLDGPPVDPEDSDNSAIYQGLNDSTLTLDLDFPQ 65
 DB 1 LGSAGRGFGNKGPGMDGPDLDGPPVDPEDSDNSAIYQGLNDSTLTLDLDFPQ 60
 QY 66 CGVVMNKRRTGQPMHIYLDKGTGPKGDATSYSDPPAKAVMPDGDQSGSLKYS 125
 DB 61 CGVVMNKRRTGQPMHIYLDKGTGPKGDATSYSDPPAKAVMPDGDQSGSLKYS 120
 QY 126 LARKKPPMNSMKGGLPPREGGMPPLRGPGGPGGPGMGMGGRGGRGFPFPGPR 185
 DB 121 LARKKPPMNSMKGGLPPREGGMPPLRGPGGPGGPGMGMGGRGGRGFPFPGPR 180
 QY 186 GSRGNPSSGGVYQHRADGWQCPNPGCGNQNFAMRTGCKQCKAPKTBGLPPPPPGGR 245
 DB 181 GSRGNPSSGGVYQHRADGWQCPNPGCGNQNFAMRTGCKQCKAPKTBGLPPPPPGGR 240
 QY 246 GRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRG 305
 DB 241 GRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRG 300
 QY 306 PLMEQMGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGR 338
 DB 301 PLMEQMGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGR 333
 RESULT 9
 Q6P3NO PRELIMINARY; PRT; 674 AA.
 ID Q6P3NO
 AC Q6P3NO
 DT 05-JUL-2004 (TEMBLrel. 27, Created)
 DT 05-JUL-2004 (TEMBLrel. 27, Last sequence update)
 DT 05-JUL-2004 (TEMBLrel. 27, Last annotation update)
 DE Hypothetical protein MGC76258.
 GN Name=MGC76258;
 OS Xenopus tropicalis (Western clawed frog) (Silurana tropicalis).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipridae; Pipidae;
 OC Xenopodinae; Xenopus.
 OK NCBI_TaxID=8364;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC TISSUE=Embryo;
 RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
 RA Strausberg R.L., Feilgold B.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shemmen C.M., Schuler G.D.,
 RA Altschuler S.F., Zeeberg B., Buettner K.H., Scheffer C.F., Boultard G.G.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Caaveira T.L., Schetz T.E.,
 RA Brownstein M.J., Ueda T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loughell N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
 RA Bosak S.A., McGowan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Boultard G.G.,
 RA Blakesley R.W., Touchman J.W., Green B.D., Dickson M.C.,
 RA Rodriguez A.C., Griemwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
 RA Krzyzanski M.I., Skalka U., Smalls D.E., Scherch A., Schein J.E.,
 RA Jones S.J., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).

[2]
 RN SEQUENCE FROM N.A.
 RC TISSUE=Embryo;
 RA Klein S., Gerhard D.S.;
 RL Submitted (DEC-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC063928; AAH63928.1; -
 DR InterPro: IPR006162; Pfam: Pfam. S.
 DR InterPro: IPR000504; RNA_rec_mot.
 DR InterPro: IPR001876; Znf_RanBP.
 DR Pfam; PF00076; RRM_1; 1.
 DR Pfam; PF00641; Zf-RanBP; 1.
 DR SMART; SMO0360; RRM; 1.
 DR SMART; SMO0547; Znf_RBZ; 1.
 DR PROSITE; PS00102; PHOSPHOPANTHETHEINE; UNKNOWN_1.
 DR PROSITE; PS01358; ZF_RANBP2_1; 1.
 DR PROSITE; PS01358; ZF_RANBP2_2; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 674 AA; 71082 MW; C4B34C7B95449B0D CRC64;
 Query Match 69.5%; Score 1336; DB 2; Length 674;
 Best Local Similarity 72.0%; Pred. No. 2,4e-69;
 Matches 250; Conservative 37; Mismatches 34; Indels 26; Gaps 11;
 QY 1 GGRGNGSAGEGFGFNKPGPMDEGPDLDGP-----PVPDPEDSDNSAIYVQGLNDV 54
 DB 345 GSRGNGSGGDRAGFGSKPGGLDDGPDLDGPMPPLPLPLDLDELESSITYVQGLNDV 404
 QY 55 TLDDLADFPKQCGVVKNNKRTGQPMTHIYLDKETGKPKDGYVSYEDPPTAKAAVWFQD 114
 DB 405 TVBEIVDFHCGDVINKRTGEBVLVLPMDKETGKPKGCTVSPEDPSAKTAIELCDG 464
 QY 115 KDFQSKLVSLARKKPPMNSMRGGLPREGRGMPPPLRGCGPGCGPGRMGGRG 174
 DB 465 KDLNGKVKVSLARKKSLGSMRGSLLDNRGCPPLR-----GGPMKRLGGRG 515
 QY 175 DRGGPPRGRGSRGNPSGGGVVQHRAGDMQCPNPGCGNQNFAWRTGNCCKAPKPDG 234
 DB 516 ERGGFMRGRGRGSPV-SGNVQHRAGDMQCPNPGCGNQNFAWRTGNCCKAPKPDG-- 572
 QY 235 P-PPPPPGDGRGPGMRGGRGGLMDRGPGCGMFRGGRGGRGGRGGRGGRG 293
 DB 573 PPPPPPGGGRGRGGR-GMRGR-GLMRGVP-GMFRGGRGGRGGR-GRMGR-GRG 627
 QY 294 GGRGPGGPPPGPLMEQMGRRG--RGPGKMDKGEHRRDRPY 338
 DB 628 GGRGPGGPPPGPLMEQMGRRIGGGRGGRGSPKMDKSEHRRDRERY 674
 RESULT 10
 ID Q6DUQ4 PRELIMINARY; PRT; 673 AA.
 AC Q6DUQ4;
 DT 25-OCT-2004 (T-EMBLrel. 28, Created)
 DT 25-OCT-2004 (T-EMBLrel. 28, Last sequence update)
 DE 25-OCT-2004 (T-EMBLrel. 28, Last annotation update)
 DE Hypothetical protein MGC76258.
 GN Name=MGC76258;
 OS Xenopus tropicalis (Western clawed frog) (Silurana tropicalis).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipridae; Pipidae;
 OC Xenopodinae; Xenopus.
 OC NCB1_TaxID=8364;
 RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
 RP TISSUE=Embryo;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Schuler G.D.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Hootch S.F., Zeeberg B., Bueltow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heien F.,
 RA Dlatchenko L., Marusik K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Schiezel T.E.,

RA Brownstein M.J., Uedín T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loggellano N.A., Peters G.U., Abramson R.D., Mullaly S.J.,
 RA Bosak S.A., McSwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahy J., Helton E., Ketterman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
 RA Krzywiński M.I., Skalek U., Smallos D.E., Schnerch A., Schein J.E.,
 RA Jones S.J., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Embryo;
 RA Klein S., Gerhard D.S.;
 RL Submitted (JUN-2004) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC075120; AAH75120.1; -
 DR GO; GO:0005634; Cnucleus; IEA.
 DR GO; GO:0003676; F-nucleic acid binding; IEA.
 DR GO; GO:0008270; F-zinc ion binding; IEA.
 DR InterPro: IPR006162; Pfam: Pfam. S.
 DR InterPro: IPR001368; RNA-binding_EWS.
 DR InterPro: IPR005004; RNA_rec_mot.
 DR InterPro: IPR001876; Znf_RanBP.
 DR Pfam; PF00076; RRM_1; 1.
 DR Pfam; PF00641; Zf-RanBP; 1.
 DR PIRSF; PIRSF002101; RNA-binding_EWS; 1.
 DR SMART; SMO0360; RRM; 1.
 DR SMART; SMO0547; Znf_RBZ; 1.
 DR PROSITE; PS00102; PHOSPHOPANTHETHEINE; UNKNOWN_1.
 DR PROSITE; PS01358; ZF_RANBP2_1; 1.
 DR PROSITE; PS01358; ZF_RANBP2_2; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 673 AA; 70995 MW; BD8A17F25991A74 CRC64;
 Query Match 68.8%; Score 1321.5; DB 2; Length 673;
 Best Local Similarity 71.8%; Pred. No. 1.6e-68;
 Matches 249; Conservative 37; Mismatches 34; Indels 27; Gaps 12;
 QY 1 GGRGNGSAGEGFGFNKPGPMDEGPDLDGP-----PVPDPEDSDNSAIYVQGLNDV 54
 DB 345 GSRGNGSGGDRAGFGSKPGGLDDGPDLDGPMPPLPLPLDLDELESSITYVQGLNDV 403
 QY 55 TLDDLADFPKQCGVVKNNKRTGQPMTHIYLDKETGKPKDGYVSYEDPPTAKAAVWFQD 114
 DB 404 TVBEIVDFHCGDVINKRTGEBVLVLPMDKETGKPKDGYVSPEDPSAKTAIELCDG 463
 QY 115 KDFQSKLVSLARKKPPMNSMRGGLPREGRGMPPPLRGCGPGCGPGRMGGRG 174
 DB 464 KDLNGKVKVSLARKKSLGSMRGSLLDNRGCPPLR-----GGPMKRLGGRG 514
 QY 175 DRGGPPRGRGSRGNPSGGGVVQHRAGDMQCPNPGCGNQNFAWRTGNCCKAPKPDG 234
 DB 515 ERGGFMRGRGRGSPV-SGNVQHRAGDMQCPNPGCGNQNFAWRTGNCCKAPKPDG-- 571
 QY 235 P-PPPPPGDGRGPGMRGGRGGLMDRGPGCGMFRGGRGGRGGRGGRGGRG 293
 DB 572 PPPPPPGGGRGRGGR-GMRGR-GLMRGVP-GMFRGGRGGRGGR-GRMGR-GRG 626
 QY 294 GGRGPGGPPPGPLMEQMGRRG--RGPGKMDKGEHRRDRPY 338
 DB 627 GGRGPGGPPPGPLMEQMGRRIGGGRGGRGSPKMDKSEHRRDRERY 673
 RESULT 11
 ID Q6NIX1 PRELIMINARY; PRT; 624 AA.
 AC Q6NIX1;
 DT 05-UTL-2004 (T-EMBLrel. 27, Created)

DT 05-JUL-2004 (TrEMBLrel. 27, last sequence update)
 DT 05-JUL-2004 (TrEMBLrel. 27, last annotation update)
 DE Wu:fc04c01 protein (Fragment).
 GN ORFNames=wu:fc04c01;
 OS Brachydanio rerio (Zebrafish) (Danio rerio).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
 OC Cyprinidae; Danio.
 NC NCBT_TaxID=7955;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Embryo;
 RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stopleton M., Soares M.B., Bonaldo M.F., Carninci P., Scheetz T.E.,
 RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loggellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
 RA Bosak S.A., McGowan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahy J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
 RA Krzywiński M.I., Skalska U., Smalins D.E., Schnerch A., Schein J.E.,
 RA Jones S.J., Matra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Embryo;
 RX Submitted (APR-2004) to the EMBL/GenBank/DBJ databases.
 RA Strausberg R.;
 RA EMBL; BC068397; AAH68397.1; -; wu:fc04c01.
 RA ZFIN; ZDB-GENE-030131-2317; wu:fc04c01.
 RA InterPro; IPR005054; RNA_rec_mot.
 RA InterPro; IPR005829; Sug_transporter.
 RA InterPro; IPR001876; Znf_RanBP.
 RA Pfam; PF00076; RRM_1; 1.1; Znf_RanBP.
 RA Pfam; PF00641; Znf_RanBP; 1.
 RA SMART; SM00360; RRM; 1.
 RA SMART; SM00547; Znf_RBZ; 1.
 RA PROSITE; PS00102; RRM; 1.
 RA PROSITE; PS00216; SUGAR_TRANSPORT_1; UNKNOWN_1.
 RA PROSITE; PS01358; ZF_RANBP2_1; 1.
 RA PROSITE; PS0199; ZF_RANBP2_2; 1.
 FT NON_TER
 SQ SEQUENCE 624 AA; 64731 MW; D0113DF02861F221 CRC64;

Query Match 61.8%; Score 1187.5; DB 2; Length 624;
 Best Local Similarity 67.9%; Pred. No. 8e-61;
 Matches 225; Conservative 33; Mismatches 42; Indels 37; Gaps 15;

DT 1 GG--RGMSGSGRGGGFGNFKGFGPMDEGPDIDGPPVDDSDNSAIVYQGLNDSTVLTDD 58
 DB 308 GGSRSRGMGAGDGRGFSKPG-----DGEWGAPERQD--DSNSTIYITGLTETNATLEER 359
 QY 59 LADPFQCGVYVKNKRTGCGMIMHITLDKNGKRGATVYEDPPTAKAVEMPDGKDPQ 118
 DB 360 VADPFQSGIIRINKRTGLPAVNIYTDKDGKRGATLSYEPPSKAAVEMPDGKDPQ 419
 QY 119 GSKLTKSLARKKPPMNSMRGGLPRREGRPPLRGSGPGGPGGPMGSG--GRGGDRG 177
 DB 420 GKTLTKSMARKKPMGMRGGM-----PMRGDGRGMGRGGMGSGGGRGGDRG 469
 QY 178 GPPPR--GPRG--SGNPSGSGNVQHRAGDMQCPNPGCGNQNPAVTECNQCKAPKPGFLP 235
 DB 470 GFMPRGGPRGMRGGRGPT--GGNMQORAGDMQCPNAGCGNQNPAVTECNQCKAPKPGFGP 528

QY 236 PEPPEGGDRGGGPGGKRGGLMDRGPGGM--PRGGRGDRGGFRGGRGMRGGRG 293
 DB 529 PPF--PPGDRGGKGGGGRGGRG--MDRGFRGPGGFRGGRVNDGFR--GGMGRGGRG 584
 QY 294 GGRGPGGPPGPMQMGRRGGRGPGKMD--KEHROERRDRY 338
 DB 585 G---RGGGP---MDMGRRGGRGPPGKMDKDHQDRRRERY 624
 RESULT 12
 ID Q7SZR6 PRELIMINARY; PRT; 623 AA.
 AC Q7SZR6;
 DT 01-OCT-2003 (TrEMBLrel. 25, Created)
 DT 01-OCT-2003 (TrEMBLrel. 25, last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 26, last annotation update)
 DE Wu:fc04c01 protein (Fragment).
 GN ORFNames=wu:fc04c01;
 OS Brachydanio rerio (Zebrafish) (Danio rerio).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
 OC Cyprinidae; Danio.
 NC NCBT_TaxID=7955;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Embryo;
 RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stopleton M., Soares M.B., Bonaldo M.F., Carninci P., Scheetz T.E.,
 RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loggellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
 RA Bosak S.A., McGowan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahy J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
 RA Krzywiński M.I., Skalska U., Smalins D.E., Schnerch A., Schein J.E.,
 RA Jones S.J., Matra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Embryo;
 RX Submitted (AUG-2003) to the EMBL/GenBank/DBJ databases.
 RA Strausberg R.;
 RA EMBL; BC056281; AAH56281.1; -;
 RA HSSP; Q9Y589; IP27.
 RA ZFIN; ZDB-GENE-030131-2317; wu:fc04c01.
 RA InterPro; IPR005054; RNA_rec_mot.
 RA InterPro; IPR005829; Sug_transporter.
 RA InterPro; IPR001876; Znf_RanBP.
 RA Pfam; PF00076; RRM_1; 1.
 RA Pfam; PF00641; Znf_RanBP; 1.
 RA SMART; SM00360; RRM; 1.
 RA SMART; SM00547; Znf_RBZ; 1.
 RA PROSITE; PS00102; RRM; 1.
 RA PROSITE; PS00216; SUGAR_TRANSPORT_1; UNKNOWN_1.
 RA PROSITE; PS01358; ZF_RANBP2_1; 1.
 RA PROSITE; PS0199; ZF_RANBP2_2; 1.
 FT NON_TER
 SQ SEQUENCE 623 AA; 64517 MW; BE47E2ED7FEB9F60 CRC64;

Query Match 60.7%; Score 1166.5; DB 2; Length 623;
 Best Local Similarity 67.8%; Pred. No. 1.3e-59;
 Matches 232; Conservative 31; Mismatches 42; Indels 37; Gaps 15;

```

QY 1 GG--RGMSAGRGGRGPNKRGPMDEGPDLDLGPVDPEDSDNSAIYVOGLNDSTLDD 58
DB 311 GGMSRGGMGAGRGGRGPKG-----DDEMGAPPEPD--DSBNSITYITGLTENATLER 362
QY 59 LADFFQCGVVMKNTGTQGMHIYLDKETGPKGATATYEDPPTAKAVERFDGKDFQ 118
DB 363 VADFFGHSGLIRNKGTGLPAVNIYTDKOTGKRGDATLSYEPSPAKAVERFDGKDFQ 422
QY 119 GSKLKYSLARKKPPMNSMRGILPPREGRMPPRLRGPGGPGGPGMGWG--GRGGRG 177
DB 423 GKCLKYSMAARKPMGMMRGGM-----PMRGDRGGMGRGGMGMMGRGGRGGRG 472
QY 178 GPRPR--GPRG--SRGNSGCGNVQHRAGDMQCPNPGCGNCFARTSCNCKAKRPGFLP 235
DB 473 GFMPRGGRGPRGMRGGRGPT--GGNMQORAGDMQCPNAGCGNCFARTSCNCKAKRPGFLP 531
QY 236 PPPPGGDRGRGGRGPGMRGGRGGLMDRGPGGM--FRGGRGGRGGRGGRGGRGGRG 293
DB 532 PPP--PGGDRGRGGRGPGMRGGRG--MDRGPGGPGGFRGGRGGRGGRGGRGGRG 587
QY 294 GGRGGRGPGGPPGLMEOMGRRGGRGGRGGRGMD--KGEHRRGR 334
DB 588 G--RGRGGRP--MDMGRRGRGMRPGKMDKGDHRRGR 623

RESULT 13
Q6NMA7 PRELIMINARY; PRT; 575 AA.
AC 06NMA7;
DT 05-JUL-2004 (T-EMBLrel. 27, Created)
DT 05-JUL-2004 (T-EMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (T-EMBLrel. 27, Last annotation update)
DB Wu:fb75g09 protein (Fragment).
GN Name:fb75g09 protein (Fragment).
OS Brachydanio rerio (Zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
NCBI_TaxId=7955;
RN 1)
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RA MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Ueda T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaney S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalek U., Smallos D.E., Schnerch A., Schein J.B.,
RA Jones S.J., Marra M.A.,
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences."
RT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN 2)
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RA Strausberg R.;
RA Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC067661; AAH67661.1; -.
DR ZFIN; ZDB-GENE-030131-1600; fua1.
DR InterPro; IPR000504; RNA_rec_mot.
DR InterPro; IPR001876; znf_RandDP.

```

```

DR Pfam; PF00076; RRM_1; 1.
DR Pfam; PF00641; zf-RanBP; 1.
DR SMART; SM00360; RRM; 1.
DR SMART; SM00547; znf_RBZ; 1.
DR PROSITE; PS50102; RRM; 1.
DR PROSITE; PS01358; zf_RanBP2_1; 1.
DR PROSITE; PS50199; zf_RanBP2_2; 1.
PT NON_TER 1
SQ SEQUENCE 575 AA; 60365 MM; A1650072377A7629 CRC64;

Query Match 55.8%; Score 1073; DB 2; Length 575;
Best Local Similarity 61.9%; Pred. No. 2,9e-54;
Matches 211; Conservative 33; Mismatches 59; Indels 38; Gaps 10;

QY 2 GRGMSAGRGGRGPNKRGPMDEGPDLDLGPVDPEDSDNSAIYVOGLNDSTLDD 61
DB 269 GRGMSAGRGGRGPNKRGPMDEGPDLDLGPVDPEDSDNSAIYVOGLNDSTLDD 326
QY 62 FFKQCGVVMKNTGTQGMHIYLDKETGPKGATATYEDPPTAKAVERFDGKDFQ 121
DB 327 FFKHTAIRINRRLNPPALNIYTDKOSGKPKGDATLSYEPAPAKAVERFDGKDFQ 386
QY 122 LKYSLARKKPPMNSMRGILPPREGRMPPRLRGPGGPGGPGMRGGRGGRGGRG 181
DB 387 LKYSMAARKPMGMRGGRGPMR-----GPRGDRGGMGRGGRGGRGGRG 430
QY 182 R--GPRGGRGPNPG--GGNVQHRAGDMQCPNPGCGNCFARTSCNCKAKRPGFL--LPP 237
DB 431 RGPGRGGMGNGGPGQGVQKRAWDCEPNAGCGNCFARTSCNCKAKRPGFLGTSPP 490
QY 238 PPPPGGDRGRGGRGGRGGRGGLMDRGPGGM--FRGGRGGRGGRGGRGGRGGRG 297
DB 491 F--SPGERGR--GGRGGRG--MDRGPGGM--RGWGDGRGGRGGRGGRGGRGGRG 543
QY 298 GGPGRGPPGLMEOMGRRGGRGGRGGRGGRGMD--KGEHRRGR 338
DB 544 GGP-----PMDGRGRGMRPGKMDKGDHRRGR 575

RESULT 14
Q803E3 PRELIMINARY; PRT; 578 AA.
AC Q803E3;
DT 01-JUN-2003 (T-EMBLrel. 24, Created)
DT 01-JUN-2003 (T-EMBLrel. 24, Last sequence update)
DT 01-MAR-2004 (T-EMBLrel. 26, Last annotation update)
DB Zgc:55864 protein.
GN Name:fb75g09 protein (Fragment).
OS Brachydanio rerio (Zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
NCBI_TaxId=7955;
RN 1)
RP SEQUENCE FROM N.A.
RC STRAIN=AB; TISSUE=Whole body;
RC MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Ueda T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaney S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalek U., Smallos D.E., Schnerch A., Schein J.B.,

```

Search completed: February 18, 2005, 15:30:31
Job time : 172 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: February 20, 2005, 16:33:24 ; Search time 3615.31 Seconds
(without alignments)
4530.143 Million cell updates/sec

Title: US-10-791-017a-2_COPY_319_656
Perfect score: 1922
Sequence: 1 GCRGCMGSGRGRGFKPKPG.....GGRGMDKGRHRRDRRY 338

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4708233 seqs, 24227607955 residues
Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0
Maximum DB seq length: 200000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODE=frame+ .p2n.model -DEV=xlp
-Q/cgna2.1/USPTO.epool.p/US10791017/runat.17022005.125807.22074/app.query.fasta_1.1358
-DB=GenEmbl -QFMT=fastlap -SUFFIX=p2n.rge -MINMATCH=0.1 -LOOPEXT=0 -LIST=45
-UNITS=bits -START=1 -END=1 -MATRIX=blomsu62 -TRANS=human40.cdi -LIST=45
-DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pct -NORH=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10791017_@CGN_1.1_7357_@runat.17022005.125807.22074 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -MARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl.1.*
1: gb_ba.*
2: gb_hcg.*
3: gb_in.*
4: gb_cm.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_ste.*
12: gb_sy.*
13: gb_un.*
14: gb_vl.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1922	100.0	1971	12 BT007796	BT007796 Synthetic
2	1922	100.0	1988	6 AX714390	AX714390 Sequence
3	1922	100.0	1988	9 AK056681	AK056681 Homo sapi
4	1922	100.0	2182	9 BC004817	BC004817 Homo sapi

5	1922	100.0	2371	6 A36460	A36460 Sequence 1
6	1922	100.0	2371	6 AR080100	AR080100 Sequence
7	1922	100.0	2330	6 CQ867364	CQ867364 Sequence
8	1922	100.0	2330	6 AX411125	AX411125 Sequence
9	1922	100.0	2330	6 HSEMS	X66839 H.sapiens E
10	1909	99.3	2026	6 CQ850483	CQ850483 Sequence
11	1909	99.3	2026	6 AK127624	AK127624 Homo sapi
12	1909	99.3	2326	6 CQ730352	CQ730352 Sequence
13	1907.5	99.2	2164	6 BC072442	BC072442 Homo sapi
14	1907.5	99.2	2189	9 AK056309	AK056309 Homo sapi
15	1907.5	99.2	2189	9 CR456430	CR456430 Homo sapi
16	1907.5	99.2	2364	9 BC011048	BC011048 Homo sapi
17	1903	99.0	2189	10 BC068236	BC068236 Mus muscu
18	1888.5	98.3	2188	6 AX305537	AX305537 Sequence
19	1888.5	98.3	2188	10 WMEMS	X79233 M.musculus
20	1888.5	98.3	2336	5 BC083960	BC083960 Xenopus 1
21	1878	97.7	7293	9 HSM80820	BX648769 Homo sapi
22	1816	94.5	2473	9 AK026270	AK026270 Homo sapi
23	1678	87.3	221444	2 AC113313	AC113313 Mus muscu
24	1678	87.3	247757	2 AC121282	AC121282 Mus muscu
25	1676	87.2	2440	5 AJ719366	AJ719366 Gallus ga
26	1661	86.4	132906	2 AC109802	AC109802 Canis fam
27	1661	86.4	163104	2 AC110672	AC110672 Canis fam
28	1632	84.9	193559	2 AC137500	AC137500 Homo sapi
29	1619	84.2	155815	2 AC018774	AC018774 Homo sapi
30	1619	84.2	180718	2 AL596087	AL596087 Human DNA
31	1619	84.2	182501	2 AC011221	AC011221 Homo sapi
32	1541	80.2	1783	6 CQ721057	CQ721057 Sequence
33	1443.5	75.1	226400	2 AC128482	AC128482 Rattus no
34	1443.5	75.1	235241	2 AC098823	AC098823 Rattus no
35	1443.5	75.1	263925	2 AC106522	AC106522 Rattus no
36	1406.5	73.2	2327	9 AK096787	AK096787 Homo sapi
37	1385	72.1	1462	6 CQ434162	CQ434162 Sequence
38	1336	69.5	2653	5 BC063928	BC063928 Xenopus t
39	1321.5	68.8	2683	5 BC075120	BC075120 Xenopus t
40	1187.5	61.8	2449	5 BC068397	BC068397 Dario rer
41	1181.5	61.5	2448	5 BC056281	BC056281 Dario rer
42	1165.5	60.6	270337	2 AC099212	AC099212 Rattus no
43	1105	57.5	221924	2 AC105889	AC105889 Rattus no
44	1105	57.5	225105	2 AC118907	AC118907 Rattus no
45	1073	55.8	2093	5 BC044518	BC044518 Dario rer

ALIGNMENTS

RESULT 1	BT007796	1971 bp	mRNA	linear	SYN 13-MAY-2003
LOCUS	BT007796				
DEFINITION	Synthetic construct Homo sapiens Ewing sarcoma breakpoint region 1				
ACCESSION	BT007796				
VERSION	BT007796.1	GI:30584430			
KEYWORDS	FLI CDNA.				
SOURCE	Synthetic construct				
ORGANISM	synthetic construct				
REFERENCE	1 (bases 1 to 1971)				
AUTHORS	Kaunine,N., Chen,X., Rolfs,A., Halleck,A., Hines,L., Eisenstein,S., Koundinya,M., Raphael,J., Moreira,D., Kelley,T., Labaer,J., Lin,Y., Phelan,M., and Farmer,A.				
TITLE	Cloning of human full-length cDNAs in BD Creator(TM) System Donor				
JOURNAL	Unpublished				
REFERENCE	2 (bases 1 to 1971)				
AUTHORS	Kaunine,N., Chen,X., Rolfs,A., Halleck,A., Hines,L., Eisenstein,S., Koundinya,M., Raphael,J., Moreira,D., Kelley,T., Labaer,J., Lin,Y., Phelan,M., and Farmer,A.				
TITLE	Direct Submission				
JOURNAL	Submitted (13-MAY-2003) BD Biosciences Clontech, 1020 East Meadow				
COMMENT	Circle, Palo Alto, CA 94303, USA This cDNA clone is a part of a collection of human full length expression clones generated by BD Biosciences Clontech and the Harvard Institute of Proteomics. Each cDNA has been cloned in two				

Percent Similarity: 100.004
 Best Local Similarity: 100.004
 Query Match: 100.004
 DB: 6
 Conservative: 0
 Mismatches: 0
 Indels: 0
 Gaps: 0

US-10-791-017a-2_COPY_319_656 (1-338) x AK714390 (1-1988)

QY 1 G1G1YARGLYGLYMETGLYSEALAGLYLUAAGLYLYPHEANLYPROMGLYLY 20
 DB 798 GAGGAGCGCGGTGAATGGGACCGCTGGAGACGAGGTGGCTTCAATGAAGCCCTGGTGA 857
 QY 21 PROMETASPLUGLYPROAEPLLEUAEPLLEUGLYPROVOLAAPPROMAEPLLEUAEPL 40
 DB 858 CCCATGGATGAAGACCAAGATCTTGATCTGAGCCCACTTGAATCAAGATGAAGACTCT 917
 QY 41 ASPLASERLALLETYRVALGINGLYLEUASAPSERVALTHIRLEUASAPLEUALA 60
 DB 918 GACACAGCTGCAATTAATGATGAAGATTAATGACAGTGAAGTGAATGATGATGGA 977
 QY 61 ASPPHAPHELYGLYGLYVALVALYMETASPLLYPHEANLYPROMGLYLY 80
 DB 978 GACTCTTTAAGAGTGGGAGTGGTGAATGAAGAAAGAACTGGGCAACCAAGATC 1037
 QY 81 HSLLETYRLEUASPLLYGLYTHIRGLYLYPROMGLYLYASPLATHTVALSERTYRGL 100
 DB 1038 CACATCTACTGACACAGAAACAGAAACCCAAAGCCAGTCCACAGTGTCTATGA 1097
 QY 101 ASPPROMPROTHALALYVALVALYGLYTHIRPHEANPLLYPHEANPLLYPHEANGLY 120
 DB 1098 GACCCACCCACCTGCAAGGCTGCGTGAATGGTGGTGAAGATTTTCAAGGAGG 1157
 QY 121 LYSLEULYVALSERLEUALAAGLYGLYBPROMETASPLLYPHEANLYPHEAN 140
 DB 1158 AAACCTTAAGTCTCCCTGTCTGCAAGAAACCTTCAATGAAGTGGGCTGTCTG 1217
 QY 141 PROPROARGGLUGLYYARGLYMETPROPROPROLEUARGGLYLYPROMGLYLY 160
 DB 1218 CACACCCCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1277
 QY 161 GLYPROGLYLYPROMETGLYARGLYMETGLYLYYARGLYLYYARGLYLYPHEAN 180
 DB 1278 GGTCTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1337
 QY 181 PROARGGLYPROARGGLYSEARGLYANPROMSEARGLYLYYANVALGINHAR 200
 DB 1338 CCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1397
 QY 201 ALAGLYASPLTRPGLINCYPROMAENPROMGLYCYEGLYASNGLYASNPHEALATPA 220
 DB 1398 GCTGAGAGCTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1457
 QY 221 GLUCYASNGLYCYELYSALAPROLYPROMGLYLYPHEANPROMPROPHENPRO 240
 DB 1458 GAGTGCAACAGGTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1517
 QY 241 PROGLYLYYASPARGLYARGLYLYPROMGLYLYYMETARGLYLYYARGLYLYLEU 260
 DB 1518 CCGGGGTGATCGTGGAGAGGTGGCCCTGGTGGATCGGGGAGAGAGGTGGCC 1577
 QY 261 METASPARGLYLYPROMGLYLYMETPHEARGGLYLYYARGLYLYYASPARGLYLY 280
 DB 1578 ATGGATCTGGTGGTGGCCCGGTGAATGTTCAAGAGGAGGAGGAGGAGGAGG 1637
 QY 281 PHEARGGLYLYYARGLYLYMETASPARGLYLYPHEARGGLYLYYARGLYLYYARGLY 300
 DB 1638 TTCCTGGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1697
 QY 301 GLYGLYLYPROMGLYLYPROMETGLYLYMETGLYLYYARGLYLYYARGLYLYY 320
 DB 1698 GGGGGGGGCGGCTGGAGCTTTGATGAAGATGGGAGGAGGAGGAGGAGGAGG 1757
 QY 321 PROGLYLYYMETASPLLYGLYLYHSLARGLYLYYARGLYLYYARGLYLYYARGLY 338

DB 1758 CCTGGAATAATGATTAAGCGAGCAGCCTGTCAGAGCGCAGAGATGGCCCTTAC 1811

RESULT 3
 AK056681
 LOCUS
 DEFINITION
 Homo sapiens cDNA FLJ32119 fis, clone PEBLM1000034, highly similar to RNA-BINDING PROTEIN EMS.

AK056681
 VERSION
 AK056681.1 GI:16552152
 KEYWORDS
 oligo capping; fis (full insert sequence).
 SOURCE
 Homo sapiens (human)
 ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

1 Ota, T., Suzuki, Y., Nishikawa, T., Otsuki, T., Sugiyama, T., Irie, R., Wakamatsu, A., Hayashi, K., Sato, H., Nagai, K., Kimura, K., Makita, H., Sekine, M., Ohtsuka, M., Nishi, T., Shibahara, T., Tanaka, T., Ishii, S., Yamamoto, J., Saito, K., Kawai, Y., Isono, Y., Nakamura, Y., Nagahara, K., Murakami, K., Yasuda, T., Iwayanagi, T., Wagatsuma, M., Shiratori, A., Sudo, H., Hosoiri, T., Kaku, Y., Kodaira, H., Kondo, H., Sugawara, M., Takahashi, M., Kanda, K., Yokoi, T., Furuya, T., Kikkawa, E., Omura, Y., Abe, K., Kamihara, K., Katsuta, N., Sato, K., Tanikawa, M., Yamazaki, M., Ninomiya, K., Iwihashi, T., Yamashita, H., Murakawa, K., Fujimori, K., Tanai, H., Kimata, M., Watanabe, M., Hirao, K., Chiba, Y., Ishida, S., Ono, Y., Takiguchi, S., Watanabe, S., Yosida, M., Hoshino, T., Nomura, J., Kanehori, K., Takahashi, F., Hara, H., Tanase, T., Nomura, Y., Togaya, S., Komai, F., Hara, R., Takeuchi, K., Arita, M., Imose, N., Mutsaers, K., Yuki, H., Oshima, A., Sasaki, N., Aotsuka, S., Yoshikawa, Y., Matsunawa, H., Ichihara, T., Shiohara, N., Sano, S., Moriya, S., Nomiyama, H., Satoh, N., Takami, S., Terashima, Y., Suzuki, O., Nakagawa, S., Senoh, A., Mizoguchi, H., Goto, Y., Shimizu, F., Wakebe, H., Hishigaki, H., Watanabe, T., Sugiyama, A., Takemoto, M., Kawakami, B., Yamazaki, M., Watanabe, K., Kumagai, A., Itakura, S., Fukuzumi, Y., Fujimori, Y., Komiyama, M., Tashiro, H., Tanigami, A., Fujiwara, T., Ono, T., Yamada, K., Fujii, Y., Ozaki, K., Hiro, M., Ohmori, Y., Kawabata, A., Hikiji, T., Koobatake, N., Inagaki, H., Ikema, Y., Okamoto, S., Ohtani, R., Kawakami, T., Noguchi, S., Itoh, T., Shigeta, K., Senba, T., Matsumura, K., Nagajima, Y., Mizuno, T., Morinaga, M., Sasaki, M., Togashi, T., Oyama, M., Hata, H., Watanabe, M., Komatsu, T., Mizushima-Sugano, J., Sato, T., Shirai, Y., Takahashi, Y., Nakagawa, K., Okumura, K., Nagase, T., Nomura, N., Kikuchi, H., Maeno, Y., Yamashita, R., Nakai, K., Yada, T., Nakamura, Y., Ohara, O., Isogai, T. and Sugano, S. Complete sequencing and characterization of 21,243 full-length human cDNAs Nat. Genet. 36 (1), 40-45 (2004)

JOURNAL

PUBMED

REFERENCE

AUTHORS

1758 CCTGGAATAATGATTAAGCGAGCAGCCTGTCAGAGCGCAGAGATGGCCCTTAC 1811
 2 Ishbaeshi, T., Kanehori, K., Yosida, M., Watanabe, S., Ishida, S., Ono, Y., Hoshino, T., Hirao, K., Murakawa, K., Takiguchi, S., Kusano, J., Watanabe, M., Fujimori, K., Tanai, H., Ishida, M., Yamashita, H., Chiba, Y., Sugiyama, T., Irie, R., Otsuki, T., Sato, H., Wakamatsu, A., Ishii, S., Yamamoto, J., Isono, Y., Kawai, H., Saito, K., Nishikawa, T., Kimura, K., Matsuo, K., Nakamura, Y., Sekine, M., Kikuchi, H., Kanda, K., Wagatsuma, M., Takahashi, F., Oshima, A., Sugiyama, A., Kawakami, B., Suzuki, Y., Sugano, S., Nagahara, K., Maeno, Y., Nagai, K. and Isogai, T.
 NEDO human cDNA sequencing project
 UNPUBLISHED
 3 (bases 1 to 1988)
 Isogai, T., Otsuki, T. and Sugiyama, T.
 Direct Submission
 Submitted (24-OCT-2001) Takao Isogai, Helix Research Institute, Genomics Laboratory, 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan (E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)

JOURNAL

AUTHORS

REFERENCE

TITLE

COMMENT

NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology (RAB); cDNA library construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: RAB and

KEYWORDS
SOURCE Homo sapiens (human).
ORGANISM Homo sapiens
REFERENCE Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
AUTHORS 1 (bases 1 to 2371)
TITLE Aurias,A., Delattre,O., Desmazaie,C., Melot,T., Peter,M.,
Nucleic Acid Corresponding to a Gene of Chromosome 22 Involved in
RECURRENT CHROMOSOMAL TRANSLOCATIONS ASSOCIATED WITH THE
DEVELOPMENT OF CANCEROUS TUMORS
JOURNAL Patent: WO 9323549-A 1 25-NOV-1993;
CENTRE NAT RECH SCIENT (FR)
COMMENT Other publication FR 2691475 931126
Other publication JP 85009647 960206.
FEATURES
Source 1. .2371
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
/dev_stage="FOETUS"

ORIGIN

Alignment Scores:
Pred. No.: 3.5e-47 Length: 2371
Score: 1922.00 Matches: 338
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: Gaps: 0

US-10-791-017a-2_COPY_319_656 (1-338) x A36460 (1-2371)

QY 1 G1G1YARGLYGLYMETGLYSEALAG1GLUARG1GLYLYPHEANLYSPROGLYGLY 20
DB 979 GGAGGAGCGGTGGAATGGGAGCGCTGGAGAGGAGTGGCTTCAATTAAGCCTGGTGA 1038

QY 21 PROMETASPG1UG1YPROASPLEUASPLEUG1YPROVALASPROMASPG1UASPSER 40
DB 1039 CCCATGATGAAGACCAAGATCTTGATCTAGAGCCCTCTTATGATCCAGTGAAGACTCT 1098

QY 41 ASPASNSER1A1LETYRVALG1NG1YLEUASNSPSEVALTHIRLEUASAPLEUALA 60
DB 1099 GACAAACAGTCAATTATGATGACAAAGATTAATGACAGTGAAGATGATGATGATGCA 1158

QY 61 ASPHETHEL1YGLYGLYVAL1VALYMETASNL1YARGTHRG1YGLINPROMET1LE 80
DB 1159 GACTTCTTAAAGACAGTGTGGGTTGTTAAGATGAACAAAGAACTGGGCAACCAATGATC 1218

QY 81 H1E1LETYRLEUASPL1YGLUTNRGL1YSPROLYSG1YASPLA1THRVALSER1YGLU 100
DB 1219 CACATCTACCTGACAAAGAAACAGAAACCCAAAGCCATGCCACAGTCTCTATGA 1278

QY 101 ASPPROB1THRAL1ALYVAL1AVAL1GLUTRPHASPL1YLYASPPHEGL1YSER 120
DB 1279 GACCCACCCCACTGCGCAAGCTGCGCGTGAATGTTTATGGAAGAAAGATTTTCAAGGAGC 1338

QY 121 LYLEULYVALSER1LEUAL1AARGLYLYSPROMET1NSER1MET1ARG1YGLYLEU 140
DB 1339 AAACCTTAAGTCTCCCTTGCTCGGAAGAACCTTCAATGAACAGTAAAGCGGGGTGCTG 1398

QY 141 PROPROARG1UG1YARGLYMET1YMET1YPROPROLEUARG1YGLYPROGLYGLY 160
DB 1399 CCAACCCGTAAGGAGCAAGACATCCACCACTCCGTGAGAGTCCAGAGAGCCCAAGCA 1458

QY 161 GLYPROGLYGLYPROMET1YARGMET1YGLYARG1YGLYASPARGLYGLY1PHEPRO 180
DB 1459 GGTCTTGAGGAGCCCATGGGTCGCGAGAGCCGTGAGAGATGAGAGAGGCTTCCCT 1518

QY 181 PROARG1YPROARG1YSEAR1YASNP1ROSE1YGLYGLYANVAL1G1NH1ARG 200
DB 1519 CCAAGAGAGACCCGGGGTTCCTCCGAGGAGAACCCCTCTGAGAGAGAAACGTCAGCA 1578

QY 201 A1AG1YASPT1RGL1NCYSPROASNP1ROGL1YCYSG1YASNG1YASNSPHEA1ATPARGTHR 220
DB 1579 GCTGAGAGACTGGACAGTCTCCCATCCGGGTGGAGAAACCAAGAACTTCGCTGAGAGCA 1638

QY 221 GLUCYASBNG1CYBLYALAPOLYSP1ROGL1UG1Y1PHELEUPROPROPH1EPROPO 240
DB 1639 GAGTGCACACATGTAAAGCCCCCAAGCCTGAAGGCTTCTCCCGCCACCTTCCGCC 1698

QY 241 PROGL1YGLYASPARGLYARGLYGLY1YMETARG1YGLY1YARG1YGLY1YARG1YGLY 260
DB 1699 CCGGTGTGATGATCGTGGACAGAGTGCCTGTGTGATGATGATGATGATGATGATGATG 1758

QY 261 METASPARGLYGLY1YPROGLYGLYMET1PHEARG1YGLY1YARG1YGLY1YASPARGLY 280
DB 1759 ATGATCGTGTGATGATCGTGGTGAATGTTCAAGAGTGGCGCGTGTGAGACAGAGTGGC 1818

QY 281 PHEARG1YGLY1YARG1YMETASPARGLYGLY1Y1PHEGLYGLY1Y1YARG1YGLY1Y 300
DB 1819 TTCCGTGTGATGATCGTGGGCAATGACCCAGAGTGTGCTTGTGTGAGAGAGAGAGAGTGGCCCT 1878

QY 301 GLYGLY1YPROGLY1YPROLEUETGL1YMETGLY1YARG1YGLY1YARG1YGLY1Y 320
DB 1879 GGGGGGCCCCCTGACCTTTGATGACAGATGGAGAGAGAGAGAGAGAGAGAGAGAGAG 1938

QY 321 PROGLY1YMETASPL1YGLYGLY1YARG1YGLY1YARG1YGLY1YARG1YGLY1Y 338
DB 1939 CTTGAAAAAATGATTAATGAGCGAGACCGCTCAGAGACGAGAGATGCGCCCTAC 1992

RESULT 6
LOCUS AR080100 2371 bp DNA linear PAT 31-AUG-2000
DEFINITION Sequence 1 from patent US 5968734.
ACCESSION AR080100
VERSION AR080100.1 GI:10006835
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 2371)
AUTHORS Aurias,A., Delattre,O., Desmazaie,C., Melot,T., Peter,M.,
Plougaetel,B., Thomas,G., and Zuchman,J.
TITLE Nucleic acid corresponding to a gene of chromosome 22 involved in
recurrent chromosomal translocations associated with the
development of cancerous tumors, and nucleic acids of fusion
resulting from said translocations
Patent: US 5968734-A 1 19-OCT-1999;
JOURNAL Location/Qualifiers
FEATURES
Source 1. .2371
/organism="unknown"
/mol_type="unassigned DNA"

ORIGIN

Alignment Scores:
Pred. No.: 3.5e-47 Length: 2371
Score: 1922.00 Matches: 338
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: Gaps: 0

US-10-791-017a-2_COPY_319_656 (1-338) x AR080100 (1-2371)

QY 1 G1G1YARGLYGLYMETGLYSEALAG1GLUARG1GLYLYPHEANLYSPROGLYGLY 20
DB 979 GGAGGAGCGGTGGAATGGGAGCGCTGGAGAGGAGTGGCTTCAATTAAGCCTGGTGA 1038

QY 21 PROMETASPG1UG1YPROASPLEUASPLEUG1YPROVALASPROMASPG1UASPSER 40
DB 1039 CCCATGATGAAGACCAAGATCTTGATCTAGAGCCCTCTTATGATCCAGTGAAGACTCT 1098

QY 41 ASPASNSER1A1LETYRVALG1NG1YLEUASNSPSEVALTHIRLEUASAPLEUALA 60
DB 1099 GACAAACAGTCAATTATGATGACAAAGATTAATGACAGTGAAGATGATGATGATGCA 1158

QY 201 A1AG1YASPTTglnCysPProAenProG1YcysG1YASnG1AAsnPha1a1TTPaTgThr 220
 Db 1598 GCTGGAGACTGGAGTGTCCCATCCGGTGTGGAACCAAGACTCCCTGGAGAAACA 1657
 QY 221 G1uCyAsnG1nCySylYsAlaPProLYSPProG1uG1YPhaLeuPProPProPhePro 240
 Db 1658 GAGTGCACCAAGTGTAAAGGCCCAAGCCCTGAAGGCTTCTCCGCCACCCCTTCCGCC 1717
 QY 241 ProG1Yg1YASpArG1YArG1Yg1YProG1Yg1YMeArG1Yg1YArG1Yg1YLeu 260
 Db 1718 CCGGGTGTGATCGTGGCAAGGTGGCCGTGTGGCATGCGGGAGAGAGAGGTGGCCCTC 1777
 QY 261 MetAspArG1Yg1YProG1Yg1YMeArPheArG1Yg1YArG1Yg1YASpArG1Yg1Y 280
 Db 1778 ATGATCGTGTGTGTCCCGGTGGAATGTTCAAGAGGTGGCCGTGTGTGAGCAAGAGGTGGC 1837
 QY 281 PheArG1Yg1YArG1YMeArPheArG1Yg1YPhaG1Yg1YArG1YArG1Yg1YPro 300
 Db 1838 TTCGTGTGTGGCCGGGAGCATGACCGAGGTGGCTTGTGTGAGAGAGAGAGGTGGCCCT 1897
 QY 301 G1Yg1YProPProG1YProLeuMeTg1uG1nMeG1Yg1YArG1Yg1YArG1Yg1Y 320
 Db 1898 GGGGGGGCCCTGGACCTTTGATGAGAACAGATGGAGAGAGAGAGAGAGAGAGAGAGAG 1957
 QY 321 ProG1YLYMeArPheArG1Yg1YUHisArG1nG1uArG1YArG1YArG1YArG1Y 338
 Db 1958 CTTGGAAAAATGATTAAGCGAGCACCGTCAGAGCGCAGAGATCGCCCTTAC 2011
 RESULT 8
 LOCUS AX411125 2390 bp DNA linear PAT 14-JUN-2002
 DEFINITION Sequence 3772 from Patent WO0229103.
 ACCESSION AX411125
 VERSION AX411125.1 GI:21443830
 KEYWORDS
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1
 AUTHORS Alvarez C., Horne D., Perez-da-Silva S. and Vockley J.G.
 TITLE Gene expression profiles in liver cancer
 JOURNAL Patent: WO 0229103-A 3772 11-APR-2002;
 GENE LOGIC INC (US)
 FEATURES
 source location/Qualifiers
 1..2390
 /organism="Homo sapiens"
 /mol_type="unassigned DNA"
 /db_xref="taxon:9606"
 /note="EMBL/GenBank Accession No. X66899"
 ORIGIN
 Alignment Scores:
 Pred. No.: 3.52e-47 Length: 2390
 Score: 1922.00 Matches: 338
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 6 Gaps: 0
 US-10-791-017a-2_copy_319_656 (1-338) x AX411125 (1-2390)
 QY 1 G1Yg1YArG1Yg1YMeArG1YSer1aG1YUArG1Yg1YPhaEnLYSPProG1Yg1Y 20
 Db 998 GGAGAGACCGGTGGAATGGGAGCGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1057
 QY 21 ProMeArPheArG1YProArPheArPheArG1YProArPheArPheArPheArPheAr 40
 Db 1058 CCGATGTGATGAGAGACCAATCTGATCTAGGCCCTCTGTGATGCCAGATGAGAGAGAGAG 1117
 QY 41 AspEnSer1a1e1YVal1G1nG1YLeuEnAspSerVal1TTPaAspArPheArPheAr 60

Db 1118 GACAACAGTGCATTTATGTACAAAGATTAAATGACAGTGTGACTTGAATGATCTGGCA 1177
 QY 61 AspPhePheLYSg1nCySylYsAlaPProLYSPProG1uG1YPhaLeuPProPProPhePro 80
 Db 1178 GACTTCTTTAAGCAGTGTGGGTGTGTAAAGTAAAGAGAGAGAGAGAGAGAGAGAGAGAG 1237
 QY 81 His1eLYLeuAspLYSg1nThx1YLYSPProLYSPProLYSPProLYSPProLYSPPro 100
 Db 1238 CACATCTACTTGCACAGAGAAACAGAGAAAGCCCAAGGGAGATGCCAGATGTCTATGAA 1297
 QY 101 AspProProThr1AlaLYsAla1aVal1G1UTTPheArPheArLYsAspPheG1nG1Ser 120
 Db 1298 GACCCACCACTGCACCAAGGCTGCGTGTGAATGTTTGAAGGAAAGATTTTCAAGGGAGAC 1357
 QY 121 Lys1eLYsVal1Ser1eLYsAla1aRg1YLYSPProPProPProPProPProPProPPro 140
 Db 1358 AAACCTTAAGTCTCTCTTGTCTGGAAGAGAGCTTCAATGAGACAGTATGCGGGGTGTCTG 1417
 QY 141 ProPProArG1uG1YArG1YMeArPProPProPProPProPProPProPProPProPPro 160
 Db 1418 CACCCCGTGAAGGAG 1477
 QY 161 G1YProG1Yg1YProMeTg1YArG1YMeTg1Yg1YArG1Yg1YASpArG1Yg1YPhaPro 180
 Db 1478 GGTCTGGGGAGCCCATGGGTGCAATGGAGGCGCTGTGAGAGATAGAGAGAGAGAGAGAG 1537
 QY 181 ProArG1YProArG1YSerArG1YLYSPProPProPProPProPProPProPProPProPPro 200
 Db 1538 CCAAG 1597
 QY 201 A1AG1YASPTTglnCysPProAenProG1YcysG1YASnG1AAsnPha1a1TTPaTgThr 220
 Db 1598 GCTGGAGACTGGAGTGTCCCATCCGGTGTGGAACCAAGACTCCCTGGAGAAACA 1657
 QY 221 G1uCyAsnG1nCySylYsAlaPProLYSPProG1uG1YPhaLeuPProPProPhePro 240
 Db 1658 GAGTGCACCAAGTGTAAAGGCCCAAGCCCTGAAGGCTTCTCCGCCACCCCTTCCGCC 1717
 QY 241 ProG1Yg1YASpArG1YArG1Yg1YProG1Yg1YMeArG1Yg1YArG1Yg1YLeu 260
 Db 1718 CCGGTGTGTGGCCGGGAGCATGACCGAGGTGGCTTGTGTGAGAGAGAGAGAGAGAGAG 1777
 QY 261 MetAspArG1Yg1YProG1Yg1YMeArPheArG1Yg1YArG1Yg1YASpArG1Yg1Y 280
 Db 1778 ATGATCGTGTGTGTCCCGGTGGAATGTTCAAGAGGTGGCCGTGTGTGAGCAAGAGGTGGC 1837
 QY 281 PheArG1Yg1YArG1YMeArPheArG1Yg1YPhaG1Yg1YArG1Yg1YArG1Yg1YPro 300
 Db 1838 TTCGTGTGTGGCCGGGAGCATGACCGAGGTGGCTTGTGTGAGAGAGAGAGAGAGAGAG 1897
 QY 301 G1Yg1YProPProG1YProLeuMeTg1uG1nMeG1Yg1YArG1Yg1YArG1Yg1Y 320
 Db 1898 GGGGGGGCCCTGGACCTTTGATGAGAACAGATGGAGAGAGAGAGAGAGAGAGAGAGAG 1957
 QY 321 ProG1YLYMeArPheArG1Yg1YUHisArG1nG1uArG1YArG1YArG1YArG1Y 338
 Db 1958 CTTGGAAAAATGATTAAGCGAGCACCGTCAGAGCGCAGAGATCGCCCTTAC 2011
 RESULT 9
 LOCUS HSEMS 2390 bp mRNA linear PRI 28-JUN-1995
 DEFINITION H.sapiens EMS mRNA.
 ACCESSION X66899
 VERSION X66899.1 GI:547565
 KEYWORDS RNA binding protein.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1
 AUTHORS Delattre, O., Zucman, J., Plougastel, B., Desmaziere, C., Melot, T.,
 Peter, M., Kovar, H., Joubert, I., de Jong, P., Rouleau, G., Auriaud, A.
 and Thomas, G.

AUTHORS Isogai, T., Yamamoto, J., Nishikawa, T., Isono, Y., Sugiyama, T., Otsuki, T., Makamatsu, A., Ishii, S., Nagai, K. and Irie, R.
 TITLE Full-length human CDNA
 JOURNAL Patent: EP 1447413-A 952 18-AUG-2004;
 Research Association for Biotechnology (JP)
 FEATURES
 source 1. .2026
 /organism="Homo sapiens"
 /mol_type="unassigned DNA"
 /db_xref="taxon:9606"
 ORIGIN
 Alignment Scores:
 Pred. No.: 7 4e-47 Length: 2026
 Score: 1909.00 Matches: 337
 Percent Similarity: 99.70% Conservative: 0
 Best Local Similarity: 99.70% Mismatches: 1
 Query Match: 99.32% Indels: 0
 Gaps: 0
 DB: 6
 US-10-791-017a-2_COPY_319_656 (1-338) x CQ850483 (1-2026)
 QY 1 G1YGYARGLYGLYMETG1YSERA1AG1Y1uarg1Y1Yphea1uYpProglY1Y 20
 DB 836 GGAAGACCGGTGGATGGGCAAGCGCTGGAAGAGAGTGGCTTCAATAGCCCTGGTGA 895
 QY 21 ProMetAspGluG1YProApPLeuApPLeuG1YProProValAspProApPLeuApPLeu 40
 DB 896 CCCATGATAGACGACGACGATCTGATCTAGCCACCTTATGATCCAGATGAGACTCT 955
 QY 41 AspAsnSerAlaIleTyrValG1NG1YLeuAsnAspSerValThrLeuAspApPLeuAla 60
 DB 956 GACACAGCTGCAATTTATGTCACAGATTTAAATACAGTGTGACTTACATGATCTGGCA 1015
 QY 61 AspPhePheLeuG1YNG1YVal1YpMetAsn1YpArgThrg1YNG1YpProme1Ile 80
 DB 1016 GACTTCTTAAAGCAGTGGGGTGTGTTAAGATGAAACAAGAACTGGGCAACCATGATC 1075
 QY 81 H1G1IeTyrLeuApPLeuApPLeuG1YpProG1YpProG1YpAspAlaThrValSerTyrGlu 100
 DB 1076 CACATCTACCTGGACAGAAACGAAAGCCCAAGCCATGCAAGTCTCTAAGAA 1135
 QY 101 AspProProThra1YpAsa1a1aValG1UTrPheAspG1YpAspPheG1NG1Yser 120
 DB 1136 GACCCACCACTGCGCAAGGCTGCGGTGATGTTTATGGAAGATTTTCAAGGAGC 1195
 QY 121 LysLeu1YpValSerLeuAlaArg1YpProPromeAsnSerMetArg1YpLeu 140
 DB 1196 AAATTAAGTCTCTTGTCTCGGAAAGCCTCAATGAACGATATGCGGGGTGCTCTG 1255
 QY 141 ProProArg1YpArg1YpMetProProPromeArg1YpProG1YpProG1YpProG1Y 160
 DB 1256 CCACCCCGTGAAGGACAGAGATCCACACACTCGTGAAGGTTCAGAGGCGCCGGA 1315
 QY 161 G1YpProG1YpPromeG1YpArgMetG1YpArg1YpArg1YpAspArgG1YpPhePro 180
 DB 1316 GGTCTGGGAGACCAAGGCTGCTGCAATGGAGGCGGTGAGAGATAGAGAGGCTTCTCT 1375
 QY 181 ProArg1YpProArg1YpSerArg1YpAsnProSerG1YpG1YpAsnValG1NH1eArg 200
 DB 1376 CCAAGAGAGACCCCGGGGTTCGCGAGGAGACCCCTCTGAGAGAGAAAGTCCAGACGGA 1435
 QY 201 AlaG1YpAspTrpG1YpProAsnProG1YpG1YpAsnG1YpAsnPheAla1YpArgThr 220
 DB 1436 GCTGGAGACCTGCAAGTATCCCAATCCGGGTGTGTAACCAAGAACTTCGCTGAGAGAA 1495
 QY 221 G1uCYaSnG1NG1YpAsa1a1aPro1YpProG1YpPheLeuPProPProPheProPro 240
 DB 1496 GAGTGAACCAAGTAAAGGCCCCCAAGCCTGAAGAGCTTCTCCGCGCAACCTTTCCGCC 1555
 QY 241 ProG1Yp1YpAspArg1YpArg1YpProG1Yp1YpMetArgG1YpArg1Yp1YpLeu 260
 DB 1556 CCGGTGTGTATCTGGCAAGAGTGGCTGTGTGATGCGGGAGAGAAAGGTGGCTCTC 1615

QY 261 MetAspArgG1Yp1YpProG1Yp1YpMetPheArgG1Yp1YpArgG1Yp1YpAspArgG1Yp 280
 DB 1616 ATGATCTGTGTGTGTCTCCGATGATTTTCAAGGCTGCGGTGAGACAGAGGTGGC 1675
 QY 281 PheArgG1Yp1YpArgG1YpMetAspArgG1YpPheG1YpG1YpArgArgG1YpPro 300
 DB 1676 TTCGGT 1735
 QY 301 G1Yp1YpProProG1YpProLeuMetG1YpMetG1Yp1YpArgArgG1Yp1YpArgG1Yp 320
 DB 1736 GGGGGGCCCCCTGAGCTTGTATGAAACAGATGGAGAGAAAGAGAGAGACGAGAGAG 1795
 QY 321 ProG1YpMetAspArgG1Yp1Yp1YpArgG1Yp1YpArgArgAspArgProTyr 338
 DB 1796 CTTGAAAATATGATTAAGGCAAGACCTTCAGAGCCGAGAGATCGGCCCTAC 1849
 RESULT 11
 AK127624
 LOCUS
 DEFINITION
 Homo sapiens CDNA FLJ45722 fib, clone HCA5M2008154, weakly similar to RNA-binding protein EMS.
 ACCESSION
 AK127624
 VERSION
 AK127624.1 GI:34534618
 KEYWORDS
 oligo capping; file (full insert sequence).
 SOURCE
 Homo sapiens (human)
 ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE
 1
 Oshima, A., Takahashi-Fujii, A., Tanase, T., Inoue, N., Takeuchi, K., Arita, M., Muraahino, K., Yuuki, H., Hara, H., Sugiyama, T., Irie, R., Otsuki, T., Sato, H., Makamatsu, A., Ishii, S., Yamamoto, J., Isono, Y., Kawai-Hio, Y., Saito, K., Nishikawa, T., Kimura, K., Yamashita, H., Matsuo, K., Nakamura, Y., Sekine, M., Kikuchi, H., Kanda, K., Wagatsuma, M., Muraikawa, K., Kanehori, K., Sugiyama, A., Kawakami, B., Suzuki, Y., Sugano, S., Nagahara, K., Masubo, Y., Nagai, K. and Isogai, T.
 NEDO human cDNA sequencing project
 JOURNAL
 Unpublished
 2 (bases 1 to 2026)
 REFERENCE
 Isogai, T. and Yamamoto, J.
 AUTHORS
 Direct Submission
 JOURNAL
 Submitted (15-JUL-2003) Takao Isogai, FLJ Project (HRI Team); 2-6-7 Kazusa-Kamatari, Kisarazu, Chiba 292-0818, Japan
 (E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)
 NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology (RAB); cDNA library construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: HRI and RAB; annotation: Reverse Proteomics Research Institute, HRI and RAB.
 FEATURES
 source 1. .2026
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="HCA5M2008154"
 /cell_type="coronary artery smooth muscle cells (HCA5MC)"
 /note="cloning vector: pM185FLJ-primary culture, coronary artery smooth muscle cells"
 ORIGIN
 Alignment Scores:
 Pred. No.: 7 4e-47 Length: 2026
 Score: 1909.00 Matches: 337
 Percent Similarity: 99.70% Conservative: 0
 Best Local Similarity: 99.70% Mismatches: 1
 Query Match: 99.32% Indels: 0

DB: 9 Gaps: 0

US-10-791-017a-2_COPY_319_656 (1-338) x AK127624 (1-2026)

QY 1 G1YGLYARGLYGLYMETGLYSEALAGLYLARGGLYLYPHEANLYSPROGLYGLY 20

DB 836 GAGAGACGGCGTGAATGGGCAAGCGCTGAGAGAGAGTGGCTTCAATAGCCCTGGTGA 895

QY 21 PROMETASPGLUGLYPROAPLEUAPLEUGLYPROPROVALIAPPROAPGLUAPSER 40

DB 896 CCATGATGATGAGACCAAGATCTTGATCTAGGCCCACTCTAGATCCAGATGAAGACTCT 955

QY 41 ASPASERIALIETRYVALINGLYLEUASNSERVALTHREUASPAAPLEUALA 60

DB 956 GACACAGTGTCAATTTATGTACAGATTAATGACAGTGTGACTGATGATGATCTGGCA 1015

QY 61 ASPPHEPHELYSGLYCYSGLYVALYLYMETASNLVSRGTTHRGLYGLINPROMETILE 80

DB 1016 GACTTCTTTAAGCAGGTGGGTTGTTAAGATGACACAGAGAACTGGGCAACCATGATC 1075

QY 81 HSIETRYLEUASPLYGLYUTHRGLYLYSPROLYSGLYASPAATHRYVALISERTYRGLU 100

DB 1076 CACATCTACCTGACCAAGAAACAGAAAGCCCAAGGCGATGCCACAGTGTCTTATGAA 1135

QY 101 ASPPROPTHRALIYVALAVALGLUTRPHASPGLYLYSAPPHREGINGLYSER 120

DB 1136 GACCCACCCCACTGCAAGGCTGCCGTGAAATGTGTGATGAGAAATTTTCAAGGAGC 1195

QY 121 LYBLEULYVALSERLEUALAARGLYLYLSPROPMETASNSERMETARGLYGLYLEU 140

DB 1196 AACTTAAAGTCTCCCTGTCTCGAAGAACCTCTCAATGAAAGTATCGGGGTGCTG 1255

QY 141 PROPROARGGLUGLYARGLYMETPROPROPROLEUARGLYGLYPROGLYGLY 160

DB 1256 CCAACCCCGTGAAGGAGGAGGATGCCACCACTCGTGAAGTCCAGAGGCCCCAGGA 1315

QY 161 GLYPROGLYGLYPROMETGLYARGMETGLYLYARGGLYLYSAPARGGLYLYLYPHEPRO 180

DB 1316 GGTCTGGGGGAGCCCATGGGTCCCATGGAGGCGGTGAGAGAAATGAGAGGCTTCCCT 1375

QY 181 PROARGLYPROARGLYSERARGLYASNPROSERGLYGLYLYSASVALGINHSARG 200

DB 1376 CCAAGAGAACCCCGGGGTTCCCAAGGAAACCTCTGAGAGAGAAAGTCCAGACCA 1435

QY 201 ALAGLYASPTTGLINCYSROAENPROGLYCYSGLYASNLNANPHEALATTPATGTH 220

DB 1436 GCTGAGAGACTGCAAGTCCCAATCCGGGTGTGGAACCAAGAACTTCCCTGAGAGAC 1495

QY 221 GLUCYASNLINCYEYVALAVALPROLYSPROGLUGLYPHELEUPROTHROPHENPRO 240

DB 1496 GAGTGCACCAAGTGAAGGCCCAAGCCCAAGGCTTCTCCGCAACCTTCCGCC 1555

QY 241 PROGLYGLYASPARGLYARGGLYGLYLYMETARGGLYGLYARGGLYGLYLYLEU 260

DB 1556 CCGGCTGGTGTGATGTCGAGAGGTCCTGTCGATGCGAGAGAGAGAGAGAGTGCCTC 1615

QY 261 METASPARGLYGLYPROGLYGLYLYMETPHEARGGLYGLYARGGLYGLYASPARGLYGLY 280

DB 1616 ATGGAATGCTGTGTGTCCTCGGTGAATGTTCAAGAGTGGCGGTGTGAGACAGAGGTGG 1675

QY 281 PHEARGGLYGLYARGGLYMETASPARGLYGLYLYPHEGLYGLYLYARGARGGLYGLYLYPRO 300

DB 1676 TTCCTGGTGGCGGGGCAATGACCAAGGTGGCTTTGTGTGAGAGAAACAGAGGTGGCCCT 1735

QY 301 GLYGLYPROPROGLYPROLEUMECGLUGLIMETGLYLYARGARGGLYGLYARGGLYGLY 320

DB 1736 GGGGGGCCCCCTGACCTTTGATGACAGATGGAGAGAGAGAGAGAGAGAGAGAGAGAG 1795

QY 321 PROGLYLYMETASPLYSGLYLYNHSARGGLYGLYLYARGARGGLYGLYLYPRO 338

DB 1796 CCTGGAATAATGATGAAGGAGAGAGAGAGAGAGAGAGAGAGAGATCGGCCCTTAC 1849

RESULT 12

CQ730352

LOCUS CQ730352 2326 bp DNA linear PAT 03-FEB-2004

DEFINITION Sequence 16286 from Patent WO02068579.

ACCESSION CQ730352

VERSION CQ730352.1 GI:42303963

KEYWORDS

SOURCE

ORGANISM

Homo sapiens (human)

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

1 Venter, C.J., Adams, M.C., Li, P.W., and Myers, E.W. kites, such as nucleic acid arrays, comprising a majority of humenexons or transcripts, for detecting expression and other uses thereof

JOURNAL Patent: WO 02068579-A 16286 06-SEP-2002;

FEATURES

source

1. .2326

/organism="Homo sapiens"

/mol_type="unassigned DNA"

/db_xref="taxon:9606"

ORIGIN

Alignment Scores:

Pred. No.: 8.11e-47 Length: 2326

Score: 1909.00 Matches: 338

Percent Similarity: 99.71% Conservative: 0

Best Local Similarity: 99.71% Mismatches: 0

Query Match: 99.32% Indels: 1

DB: 6 Gaps: 0

US-10-791-017a-2_COPY_319_656 (1-338) x CQ730352 (1-2326)

QY 1 G1YGLYARGLYGLYMETGLYSEALAGLYLARGGLYLYPHEANLYSPROGLYGLY 20

DB 950 GAGAGACGGCGTGAATGGGCAAGCGCTGAGAGAGAGTGGCTTCAATAGCCCTGGTGA 1009

QY 21 PROMETASPGLUGLYPROAPLEUAPLEUGLYPROPROVALIAPPROAPGLUAPSER 40

DB 1010 CCATGATGATGAGACCAAGATCTTGATCTAGGCCCACTGATATCAAGTGAAGACTCT 1069

QY 41 ASPASERIALIETRYVALINGLYLEUASNSERVALTHREUASPAAPLEUALA 60

DB 1070 GACACAGTGTCAATTTATGTACAGATTAATGACAGTGTGACTCTAGATGATCTGGCA 1129

QY 61 ASPPHEPHELYSGLYCYSGLYVALYLYMETASNLVSRGTTHRGLYGLINPROMETILE 80

DB 1130 GACTTCTTTAAGCAGGTGGGTTGTTAAGATGACAGAGAACTGGGCAACCATGATC 1189

QY 81 HSIETRYLEUASPLYGLYUTHRGLYLYSAPPHREGINGLYSER 100

DB 1190 CACATCTACCTGACCAAGAGAAACAGAAAGCCCAAGGAGATGCCACAGTGTCTATGAA 1249

QY 101 ASPPROPTHRALIYVALAVALGLUTRPHASPGLYLYSAPPHREGINGLYSER 120

DB 1250 GACCCACCCCACTGCAAGGCTGCCGTGAAATGTTGAGGAGAAATTTTCAAGGAGC 1309

QY 121 LYBLEULYVALSERLEUALAARGLYLYLSPROPMETASNSERMETARGGLYGLYLEU 140

DB 1310 AACTTAAAGTCTCCCTGTCTGAGAGAGCTTCCATGAAAGATGCGGGGTGCTG 1369

QY 141 PROPROARGGLUGLYARGLYMETPROPROPROLEUARGLYGLYPROGLYGLY 160

DB 1370 CCAACCCCGTGAAGGAGGAGGATGCCACCACTCGTGAAGTCCAGAGGCCCAAGGA 1429

QY 161 GLYPROGLYGLYPROMETGLYARGMETGLYLYARGGLYGLYASPARGLYGLYLYPHEPRO 180

DB 1430 GGTCTGGGGGAGCCCATGGGTGCAATGGAGGCGGTGAGAGAGATGAGAGAGGCTTCCCT 1489

QY 181 PROARGLYPROARGLYSERARGLYASNPROSERGLYGLYLYSASVALGINHSARG 200

DB 1490 CCAAGAGAACCCCGGGGTTCCGAGGAAACCTCTTGAAGAGAGAAAGTCCAGACCA 1549

PGGPMDEGPDLDGPPVDPDSDNSAIYVQGLNDSTVTLDLADPFKCGCVVMKRT
GOMHITVLDKETGKPKGDA TVSEYEDPTAKAVEMWDSGDFGSGSLKSLAKKPPM
NSMRGGLPRREGKMPPLRGPGGPGGPGGPGGPGGPGGPGGPGGPGGPGGPGG
SGGAVQHRAGDMQCIPNPGGNGQFAARTCNQCKAPKPGFLPPPPPGGGRGNGP
PGGMRGGRGLMDRGPGGMPFRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRG
MEQMGRRGGRGGRGPGKMDKGEHROERRDRRY"

ORIGIN

Alignment Scores:

Pred. No.:	8.59e-47	Length:	2189
Score:	1907.50	Matches:	337
Percent Similarity:	99.70%	Conservative:	0
Best Local Similarity:	99.70%	Mismatches:	0
Query Match:	99.25%	Indels:	1
DB:	9	Gaps:	1

US-10-791-017a-2_copy_319_656 (1-338) x CR456490 (1-2189)

```
QY 1 G1YGIYARGLYGLYMETGLYSEALAG1YGLUARG1YGL1YPHEANLYSPROG1YGLY 20
DB 988 GGAGAGCGCGGTGGATGGC--GCTGGAGAGCGAGTGCTTCAATTAAGCCTGTGGA 1044
QY 21 ProMetAspGluG1YProAspLeuAspLeuG1YProProValAspProAspGluAspSer 40
DB 1045 CCCATGGATGAAGGACCAAGATCTTGATCTAGGCCCACTGTAGATCCAGATGAAGACTCT 1104
QY 41 AspAsnSerAlaIleTyValG1NG1YLeuAsnAspSerValThreLeuAspAspLeuAla 60
DB 1105 GACAAACAGTCGAATTATGTACAAGATTAAATGACAGTGTGACTCTAGATGATCTGGCA 1164
QY 61 AspPhePheLeuG1NG1YVal1YVal1YMetAsnLyAspG1YGL1YProMetIle 80
DB 1165 GACTCTTTAAGCAGTGGGGTGTTAAGATGAACAAGAACTGGGCAACCCATGATC 1224
QY 81 HisIleTyLeuAspLyG1UThrG1YLyPProLyG1YAspAlaThrValSerTyG1U 100
DB 1225 CACATCTACCTGGACAAGAAACAGAAACCCAAAGCGATGCCACAGTGTCTATGAA 1284
QY 101 AspProProThrAlaLyValAlaValG1UTrPheAspG1YLyAspPheG1NG1YSer 120
DB 1285 GACCCACCCCACTCCAAAGGCTGCCGTGGAAATGGTTGATGGAAAGATTTCAGAGGAGC 1344
QY 121 LyLeuLyValSerLeuAlaArgLyValPProProMetAsnSerMetArgG1YGL1YLeu 140
DB 1345 AAACCTTAAGTCTCCCTTGCTCGAAGAAAGCTTCAATGAACAGTATGCCGGGTGTCTG 1404
QY 141 ProProArgG1UArg1YMetProProProLeuArgG1YGL1YProG1YGL1YProG1Y 160
DB 1405 CCACCCCGTGAAGGCGAGGCAATGCCACCACTCCGTGAGGTCCAGAGGCCCAAGGA 1464
QY 161 G1YProG1YGL1YProMetG1YArgMetG1YArgG1YGL1YAspArgG1YGL1YPhePro 180
DB 1465 GGTCTCGGGGAGCCCATGGGTCCCATGGGAGGCCGTGAGAGATAGAGGAGGCTTCCCT 1524
QY 181 ProArgG1YProArgG1YSerArgG1YAsnProSerG1YGL1YAsnValG1NH1SerG 200
DB 1525 CCAAGAGAACCCCGGGGTTCCTCGAAGGAAACCTCTGAGAGAGAAACGTCAGACCCGA 1584
QY 201 AlaG1YAspTrpG1NG1YProAsnProG1YCyG1YAsnG1NAsnPheAlaTPArgThr 220
DB 1585 GCTGGAGACTGGCAAGTCCCATCCGGGTGTGAAACCAAGAACTTCGCTGGAGAAACA 1644
QY 221 G1UCyAsnG1NCyAlYbAlaProLyAspProG1UG1YPheLeuProProPheProPro 240
DB 1645 GAGTGCAACCAAGTGAAGGCCCAAAAGCTGAAGGCTTCTCCGCGCACCCCTTCGCGCC 1704
QY 241 ProG1YGL1YAspArgG1YArgG1YGL1YProG1YGL1YMetArgG1YGL1YArgG1YGL1YLeu 260
DB 1705 CCGGGTGTGTGTGTGGCAAGGTGGCTCTGTGTGCATGCGGGAGAGAGGTGGCTTC 1764
QY 261 MetAspArgG1YGL1YProG1YGL1YMetPheArgG1YGL1YArgG1YGL1YAspArgG1YGL1Y 280
DB 1765 ATGGATCTGTGTGTCTCCGCTGGAAATTTCAAGAGTGTGCGGTGTGTGAGACAGAGGTGGC 1824
```

```
QY 281 PheArgG1YGL1YArgG1YMetAspArgG1YGL1YPheG1YGL1YArgArgG1YGL1YPro 300
DB 1825 TTCCGTGTGTGGCCGGGCAATGAGCCAGAGTGTGCTTTGTGTGAAGAAAGAGAGGAGGCTT 1884
QY 301 G1YGL1YProProG1YProLeuMetG1UG1NMetG1YGL1YArgArgG1YGL1YArgG1YGL1Y 320
DB 1885 GGGGGGCCCCCTGACCTTGATGAAACAGATGGAGAAAGAAAGAGAGGAGGAGGAGGAGGAG 1944
QY 321 ProG1YLyMetAspLyG1YGL1NH1SerG1NG1YArgArgG1YAspArgProTy 338
DB 1945 CTTGAAAAATGATTAAGCGAGACCCGTCAAGAGCGCAGAGATGGCCCTTAC 1998
```

Search completed: February 21, 2005, 02:07:41
Job time : 3639.31 secs

Large Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: February 20, 2005, 16:18:03 ; Search time 427.771 Seconds
(without alignments)
4677.443 Million cell updates/sec

Title: US-10-791-017a-2_COPY_319_656

Perfect score: 1922

Sequence: 1 GGRGGMGSGRGGFRKPKPG.....GGRGMDKGRHQRDRRY 338

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODE=frame+.p2n.model -DEV=xlp
-Q/cgn2.1/USPTO.spool.p/US10791017/runat.17022005.125806.22064/app.query.fasta_1.1358
-DB=N Geneseg.16Dec04 -QFMT=fastcap -SUFPRX=p2n.rng -MINMATCH=0.1 -LOOPEXT=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=humand0.cdt
-LIST=45 -DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pct -NORTEXT -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10791017@cgn.1.1.886 @runat.17022005.125806.22064 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -MARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

N_Geneseg.16Dec04:*
1: geneseg.19808:*
2: geneseg.19908:*
3: geneseg.20008:*
4: geneseg.20018:*
5: geneseg.20018:*
6: geneseg.20028:*
7: geneseg.20028:*
8: geneseg.20038:*
9: geneseg.20038:*
10: geneseg.20038:*
11: geneseg.20038:*
12: geneseg.20048:*
13: geneseg.20048:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1922	100.0	1988	10	ADA53506	ADA53506 Human ccd
2	1922	100.0	2176	6	AA62262	AA62262 cDNA sequ
3	1922	100.0	2371	2	AA050643	AA050643 Human Ewe
4	1922	100.0	2372	13	AD086032	AD086032 Human tum
5	1922	100.0	2388	5	AA570647	AA570647 DNA encod

5	1922	100.0	2390	6	ABR64628	ABR64628 Human cDN
7	1922	100.0	2390	6	ABR64822	ABR64822 Human ben
8	1922	100.0	2390	6	ABN97274	ABN97274 Gene #377
9	1922	100.0	2390	13	AD516277	AD516277 Human cDN
10	1922	100.0	2390	13	ACN40903	ACN40903 Tumour-as
11	1909	99.3	2026	13	AD074466	AD074466 Full leng
12	1888.5	98.3	2188	6	AB199383	AB199383 Mouse isc
13	1810	94.2	1807	13	ADP56333	ADP56333 Human PRO
14	1777	92.5	2273	6	AA562623	AA562623 cDNA sequ
15	1742	90.6	3309	2	AA050646	AA050646 Human Ewe
16	1619	84.2	2177	13	ACN37537	ACN37537 Tumour-as
17	1385	72.1	1462	4	AA126725	AA126725 Human bre
18	1181	61.4	2299	11	ACN86697	ACN86697 Breast ca
19	961	50.0	550	9	ACH40478	ACH40478 Human foe
20	878	45.7	1822	6	AB573275	AB573275 DNA encod
21	878	45.7	1822	13	AD146468	AD146468 Human NF-
22	878	45.7	1824	6	AB573272	AB573272 DNA encod
23	878	45.7	1824	6	ABR64149	ABR64149 Human cDN
24	878	45.7	1824	12	AD126116	AD126116 Human cDN
25	878	45.7	1824	12	AD162788	AD162788 Human cDN
26	878	45.7	1824	13	ACN39577	ACN39577 Tumour-as
27	878	45.7	1939	12	AD577157	AD577157 Human cDN
28	878	45.7	2179	4	AAK51488	AAK51488 Human pol
29	878	45.7	2209	4	AAK51489	AAK51489 Human pol
30	878	45.7	2213	4	AAK52472	AAK52472 Human pol
31	878	45.7	2213	4	AAK52473	AAK52473 Human pol
32	873.5	42.1	1832	12	AD126112	AD126112 Human cDN
33	809.5	42.1	568	9	ACH22345	ACH22345 Human adu
34	750	39.0	2144	12	AD126206	AD126206 Human cDN
35	750	39.0	2153	6	ABR64369	ABR64369 Human cDN
36	750	39.0	2153	12	AD126208	AD126208 Human cDN
37	749	39.0	2126	3	AA64661	AA64661 cDNA encod
38	733	38.1	471	2	AAV86051	AAV86051 EST clone
39	718.5	37.4	545	3	AACT8275	AACT8275 Human can
40	703	36.6	1610	11	AD12631	AD12631 Human dis
41	622	32.4	665	13	AD054544	AD054544 Novel can
42	616.5	32.1	1619	4	AB106387	AB106387 Drosophila
43	593	30.9	361	8	ABX38686	ABX38686 Bovine ES
44	572.5	29.8	550	6	AB199315	AB199315 Mouse isc
45	530.5	27.6	6664	4	AB106386	AB106386 Drosophila

ALIGNMENTS

RESULT 1	ADA53506	ADA53506 standard; cDNA, 1988 BP.
ID	ADA53506	
AC	ADA53506	
XX		
XX		
DT	20-NOV-2003	(first entry)
XX		
DE	Human coding sequence, SEQ ID 1074.	
XX		
KW	Cytosolic; Anti-inflammatory; Osteopontin; Neuroprotective; Neotrophic;	
KW	Gene Therapy; human; secretory protein; membrane proteins; cancer;	
KW	Inflammatory disease; osteoporosis; neurological disease; gene; ss.	
XX		
OS	Homo sapiens.	
XX		
PN	EP1293569-A2.	
XX		
PD	19-MAR-2003.	
XX		
XX		
PF	21-MAR-2002; 2002EP-00006586.	
XX		
PR	14-SEP-2001; 2001JP-00328381.	
XX		
PR	24-JUN-2002; 2002US-0350435P.	
XX		
PA	(HELI-) HELIX RES INST.	
XX		
XX	(REAS-) RES ASSOC BIOTECHNOLOGY.	
PI	Isoagi T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;	

PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;
PI Seki N, Yoshikawa T, Otsuka M, Nagahara K, Masuo Y;
XX MPI; 2003-395539/38.
DR P-PSDB; ADA55145.
XX
PT New polynucleotides encoding full-length polypeptides, e.g. secretory
PT and/or membrane proteins, useful for developing medicines for diseases in
PT which the gene is involved, or as target molecules for gene therapy.
XX
PS Claim 1; SEQ ID NO 1074; 205pp; English.
XX
CC The present invention relates to novel human secretory or membrane
CC proteins (ADA54072-ADA55710) and their coding sequences (ADA52433-
CC ADA54071). The coding sequences are useful in the gene therapy of
CC diseases caused by abnormalities of the proteins, e.g. cancer,
CC inflammatory diseases, osteoporosis or neurological disease.
XX
SQ Sequence 1988 BP; 519 A; 504 C; 584 G; 381 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 1,896-60 Length: 1988
Score: 1922.00 Matches: 338
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 10 Gaps: 0

US-10-791-017a-2_COPY_319_656 (1-338) x ADA53506 (1-1988)

QY 1 G1YGLYARGLYGLYMETGLYSERIALGLYLAIRGLYGLYPHEANLYSPROGLYGLY 20
DB 798 GAGAGAGCGGTGATGGCAGGCGCTGGAGAGCGAGTCTTCAATTAAGCTGTGGA 857

QY 21 PROMETASPLUGLYPROASPYLEUASPLEUGLYPROVALASPROMASPGLUASPSER 40
DB 858 CCATGATGATGAGAGCAGATCTTGAATCTAGGCCACCTGTAGATCCAGATGAAGACTCT 917

QY 41 ASPANSERALAIETRYVALINGLYLEUANASPSERVALTHLEUASPLAUEUA 60
DB 918 GACAAACAGTGCAATTAATGACAGATTAATGACAGTGTGACTGATGCTGGCA 977

QY 61 ASPPHEPHELYSGLYNGLYVALYALYMETASPLRYTHRGYGLYINPROMETILE 80
DB 978 GACTTCTTAAAGAGTGTGGGTGTTTAAGATGAACAAGAACTGGCAACCATGATC 1037

QY 81 HIAILETYLEUASPLRYTHRGYLYSPROLYGLYASPLATHRYVALSETYRGY 100
DB 1038 CACATCTACCTGACAAAGAAACAGAAAGCCCAAGGCGATGCCAGTGTCTATGAA 1097

QY 101 ASPPROPTHRYALYALYALYALYALYALYALYALYALYALYALYALYALYALY 120
DB 1098 GACCCACCCACTGCGCAAGGCTGCGGTGAAATGTTGATGGAAAGATTTTCAAGGAGC 1157

QY 121 LYLEUAYVALSERLEUAIALYGLYLYSPROPROMETASPSERMETARGYGLYLEU 140
DB 1158 AAACCTTAAGTCTCTCTGCTCGAAGAGCTCTCAATGAACGATATGCGGGTGTCTG 1217

QY 141 PROPROARGGLYGLYARGYMETPROPROPROLEUARGYGLYPROGLYGLY 160
DB 1218 CCAACCCGCTGACAGAGGAGCATGCGCACCACTCGGTGAGAGTCCAGAGGCCACGGA 1277

QY 161 GLYPROGLYGLYPROMETGLYARGMETGLYGLYARGYGLYGLYASPLARGYGLYPHEPRO 180
DB 1278 GGTCTCGGGGAGCCCATGCTGCGATGCGAGGCGGTGAGAGATGAAGAGGCTTCCCT 1337

QY 181 PROARGLYPROARGLYSERARGLYASPSERSEGLYGLYGLYASPLVALGINHARG 200
DB 1338 CCAAGAGAGCCCGGGGTTCCCAAGGAGAACCTCTTGAGAGAGAGAAACGTCACACCGA 1397

QY 201 ALAGLYASPLRPGYCYSPROASMPROGLYCYGLYASPLINASPLAATPARGYTH 220
DB 1398 GCTGAGAGACTGGAGTGTCCCAATCCGGGTGTGGAAACAGAACTTCCGCTGAGAGACA 1457

QY 221 GLUCYASANGLYCYLYSEALAPROLYSPROGLYGLYPHELEUPROPROPHOPROPRO 240
DB 1458 GAGTGCACACAGATGTAAGGCCCAAGACCTGAAGGCTTCTCCGACACCTTCCGCC 1517

QY 241 PROGLYGLYASPARGLYARGYGLYPROGLYGLYMETARGYGLYARGYGLYLEU 260
DB 1518 CCGGTGTGATCTGTGCGAGAGTGGCTCTGTGCGATCGGGAGAGAGAGTGGCTC 1577

QY 261 METAPARGLYGLYPROGLYGLYMETPHEARGLYGLYARGYGLYASPLARGYGLY 280
DB 1578 ATGATCTGTGTGTCTCCCGGTGAATGTTCAAGAGGTGGCGGTGTGAGACAGAGGTGGC 1637

QY 281 PHEARGLYGLYARGYGLYMETASPARGLYGLYPHEGLYGLYGLYARGYGLYPRO 300
DB 1638 TTCCTGTGTGTGCGGGGCGATGAGCCAGGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1697

QY 301 GLYGLYPROPROGLYPROLEUMETGLYINMETGLYGLYARGYGLYGLYARGYGLY 320
DB 1698 GGGGGGCCCCCTGAGACTTGTATGAAACAGATGGAGAGAGAGAGAGAGAGAGAGAG 1757

QY 321 PROGLYLYMETASPLRYGLYGLYHIAIRGLYGLYARGYGLYARGYGLYARGYGLY 338
DB 1758 CCTGGAATAATGATTAAGCGAGCAGCTCAGAGCGAGAGATGCGCCCTAC 1811

RESULT 2
ID AAS62262/C
ID AAS62262 standard; cDNA, 2176 BP.
XX
XX AAS62262;
XX
XX 14-FEB-2002 (first entry)
XX
XX cDNA sequence #49 encoding novel human secreted protein.
XX
XX Human secreted protein; hyperproliferative disorder; autoimmune disorder;
XX immune deficiency disorder; blood disorder; inflammatory disorder;
XX infectious disorder; gene therapy; antimicrobial; hepatotropic;
XX immunosuppressive; antineumatic; 88.
XX
XX Homo sapiens.
XX
XX MO200177291-A2.
XX
XX 18-OCT-2001.
XX
XX 29-MAR-2001; 2001WO-US010485.
XX
XX 06-APR-2000; 2000US-0195604P.
XX
XX (GENY) GENETICS INST INC.
XX
XX Wong GG, Clark HF, Fechtel K, Agostino MJ, Howes SH, Resnick RJ;
XX Gulukota K, Graham JR;
XX MPI; 2002-010900/01.
XX
XX New polynucleotides encoding secreted proteins useful for treating e.g.
XX asthma, HIV and Crohn's disease.
XX
XX Claim 1; Page 100; 391pp; English.
XX
XX The present invention relates to the isolation of novel cDNA sequences
XX which encode human secreted proteins. The cDNA sequences have been
XX derived from a variety of human tissues. The invention also provides a
XX method for producing proteins from these polynucleotide sequences. The
XX proteins are useful for identifying compounds that modulate their
XX activity and production, and the cell is also useful for identifying
XX compounds that modulate expression of the polynucleotide sequences
XX encoding the secreted proteins. The sequences of the invention are useful
XX for treating diseases such as hyperproliferative disorders (e.g. cancer),
XX immune deficiency disorders (e.g. severe combined immunodeficiency
XX (SCID)), autoimmune disorders (e.g. multiple sclerosis), blood disorders

(e.g. thrombocytopenia), inflammatory disorders (e.g. arthritis) and CC infectious disorders (e.g. hepatitis). The polynucleotide sequences of CC the invention are also useful in gene therapy. AAS62214-AAS62838 CC represent the cDNA sequences of the invention that encode for novel human CC secreted proteins XX

SQ Sequence 2176 BP; 413 A; 617 C; 563 G; 583 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	2.01e-60	Length:	2176
Score:	1922.00	Matches:	338
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	6	Gaps:	0

US-10-791-017a-2_COPY_319_656 (1-338) x AAS62262 (1-2176)

QY 1 G1VGIYARGLYGLYMETGLYSEALAGLYLUARGLYGLYPHEANLYSPROGLYGLY 20
 DB 1207 GAGGAGCGCGGTGAATGGGACCGCTGGAGCGAGGTGGCTTCAATAAAGCTGGTGA 1148
 QY 21 PrometaspogluniglyproaspbleuasplleuglyproprovalaspProaspgluaaspr 40
 DB 1147 CCATGATGAAGAGACCAAGTCTTGATCTAGGCCCACTGTATCCAGATGAAGACTCT 1088
 QY 41 AspAsnSerAlaIleTyrValGlnGlyLeuAsnAspSerValThreLeuAspLeuAla 60
 DB 1087 GACAAACAGTCATTTATGTACAAAGATTAAATGACAGTGTGACTAGATGATCTGGCA 1028
 QY 61 AspPhePheLeuGlnGlyValValValMetAsnLysArgThrGlyGlnPromeIle 80
 DB 1027 GAATCTTTAAAGAGTGGGTGTGTTAAGATTAACAAGAACTGGCAACCATGATC 968
 QY 81 HisIleTyrLeuAspLysGluThrGlyLysProGlyValAspAlaThrValSerTyrGlu 100
 DB 967 CACATCTACTGACCAAGAAACAGAAAGCCAAAGGCATCCACAGTGTCTTGA 908
 QY 101 AspProProThrAlaLysAlaAlaValGluTyrPheAspGlyLysAspPheGlnGlySer 120
 DB 907 GACCCACCACTGCCAAGGCTGCCGTGAGATGTTGATGGAAAGATTTCAAGGAGAC 848
 QY 121 LysLeuLysValSerLeuAlaArgLysLysProPromeAsnSerMetArgGlyGlyLeu 140
 DB 847 AAACCTTAAGTCTCCCTGCTCGAAAGAACCTTCATGAACGATAGCGGTGTCTG 788
 QY 141 ProProArgGluGlyLysArgLysMetProProProLeuArgGlyGlyProGlyGly 160
 DB 787 CCAACCCCGTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 728
 QY 161 GlyProGlyGlyProMetGlyLysArgGlyGlyLysArgGlyGlyLysArgGlyGlyPhePro 180
 DB 727 GGTCTCGGGGAGCCCATGGGTCCGATGGAGGCGGTGAGAGATAGAGAGGCTTCCT 668
 QY 181 ProArgGlyLysProArgGlyLysArgGlyLysArgProSerGlyGlyLysAsnValGlnHisArg 200
 DB 667 CCAAGAGAGCCCGGGGTCTCCGAGGAGAACCTTCGAGAGAGAGAGAGAGAGAGAGAG 608
 QY 201 AlaGlyAspTrogInCysProAsnProGlyCysGlyAsnGlnAsnPheAlaTTPArgThr 220
 DB 607 GCTGGAGATCGGAGATGCTCCCATCCGGGTGTGAGAAACAGAACTTCGCTGAGAGAG 548
 QY 221 G1VYAsnGlnCysAlaLysAlaProLysProGlyGlyPheLeuProProPheProPro 240
 DB 547 GAGTGCACACAGGTAAAGGCCCAAGACCTGAGAGGCTTCCCTCCGACACCTTTCCGCC 488
 QY 241 ProGlyGlyLysAspArgGlyLysArgGlyGlyProGlyGlyLysMetArgGlyLysArgGlyGlyLeu 260
 DB 487 CCGGT 428
 QY 261 MetAspArgGlyGlyProGlyGlyMetPheArgGlyGlyLysArgGlyGlyLysArgGlyGly 280
 DB 427 ATGGATCGT 368

QY 281 PheArgGlyGlyLysArgGlyLysMetAspArgGlyGlyPheGlyGlyLysArgGlyGlyPro 300
 DB 367 TTCCTGT 308
 QY 301 G1VGIYProProGlyProLeuMetGluGlnMetGlyGlyLysArgGlyGlyLysArgGlyGly 320
 DB 307 GGGGGGCCCCCTGAGCTTTGATGAGAACAGATGGAGAGAGAGAGAGAGAGAGAGAGAG 248
 QY 321 ProGlyLysMetAspLysGlyGlnHisArgGlnGlnLysArgAspArgProTyr 338
 DB 247 CCGGAAAATGATTAAGCGAGACCGCTCAGAGAGCGAGAGATGGCCCTTAC 194

RESULT 3

AAQ50643 standard; cDNA; 2371 BP.

AC AAQ50643;
 XX
 DT 25-MAR-2003 (revised)
 DT 26-MAY-1994 (first entry)
 XX
 DE Human Ews gene clone BFLAC5 from foetal brain cDNA library.
 KW chromosomal translocation; chimeric; chimaeric; Ewing sarcoma; Ews gene;
 KW malignant melanoma; hum-Flt-1;
 KW primitive periphereal neuroectodermal tumour; human chromosome 11;
 KW human chromosome 22; ss.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT CDS 25..1995
 FT FT /tag= a
 FT FT /product= "EWS protein"
 FT FT /transl_except= pos:1729..1731; aa:Val
 FT polyA_signal 2143..2148
 FT polyA_signal 2331..2336
 FT FT /tag= b
 FT FT /tag= c
 PN MO9323549-A2.
 XX
 XX 25-NOV-1993.
 PD
 XX
 PF 19-MAY-1993; 93WO-FR000494.
 XX
 PR 20-MAY-1992; 92FR-00006123.
 XX
 PA (CNRS) CNRS CENT NAT RECH SCI.
 PI Aurias A, Delattre O, Desmaza C, Melot T, Peter M, Plougastel B;
 PI Thomas G, Zucman J;
 XX
 DR MPI, 1993-386580/48.
 DR P-PSDB; AAR44555.
 XX
 PT New nucleic acid of EWS gene and its hybrid(s) - contg. gene sequence
 PT involved in chromosomal trans-location, also derived mRNA, probes, fusion
 PT proteins etc., for diagnosis and treatment of Ewing sarcoma and melanoma.
 XX
 PS Disclosure; Fig 6; 123pp; French.
 XX
 CC The probes 22R3 and 22R12 were used to screen a human foetal brain cDNA
 CC library (Stratagene cat. # 936206). The clone BFLAC5 was identified and
 CC sequenced. It represents the entire coding region and 3'-UTR of the Ews
 CC gene. (updated on 25-MAR-2003 to correct PN field.)
 XX
 SQ Sequence 2371 BP; 639 A; 587 C; 659 G; 486 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	2.13e-60	Length:	2371
Score:	1922.00	Matches:	338

Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 2 Gaps: 0

US-10-791-017a-2_copy_319_656 (1-338) x AAQ50643 (1-2371)

QY 1 G1YGLVARG1YGLVMECTG1YSEKALAG1YGLV1YVHEANLYSPROG1YGLY 20
 DB 979 GGAGAGACGGGTGAAAGGCGACGCTGGAGAGCGAGTGGCTTCAATTAAGCCCTGGTGA 1038
 QY 21 PROMEASPG1UG1YPROAPLEUAPLEUG1YPROPROVAL1APPROAG1UAPSER 40
 DB 1039 CCCATGATGAAGAACCAAGATCTTGATCTAGGCCCTCTTGAATCCAGATGAAGACTCT 1098
 QY 41 AAPASNSER1A1ETRYVAL1G1NG1YLEUASAPSERVAL1TH1EUAAPLEU1A 60
 DB 1099 GACAAACGTCGAATTTATGTATCAAGATTAATGACAGTGTGACTTATGATGATCGCA 1158
 QY 61 AAPHPEHEL1YSG1YVAG1YVAL1YVMECTAS1YVARG1YGL1NPROME11E 80
 DB 1159 GACTTCTTAAAGCAGTGTGGGTGTTAAAGATGAAGAACTGGGCAACCATGATC 1218
 QY 81 H1E11ETRYEUBAP1YGLUT1YVAG1YVPRO1YVAG1YVAP1A1TH1E1YRG1U 100
 DB 1219 CACATCTACCTGACCAAGAAACAGAAACCCAAAGCCATGCCAGTCTCTATGA 1278
 QY 101 AAPPROBOTH1A1ALY1A1A1A1A1A1G1UT1YVHEAP1Y1YVAP1PHEG1NG1Y 120
 DB 1279 GACCCACCCACTGCGCAAGCGCTGGTGAATGGTTTATGGAAAGATTTTCAAGGGAGC 1338
 QY 121 LY1E1U1YV1A1SER1EUA1A1ARG1YV1S1YV1APROB1E1A1N1S1E1R1E1A1R1G1Y 140
 DB 1339 AAAC1TAAATCTCTCCCTTGCTCGGAAAGCCTTCAATGAACAGTATGCGGGGTGCTG 1398
 QY 141 PROPROARG1UG1YARG1YME1YV1YVPRO1YVPRO1YVAG1YGL1YPROG1Y 160
 DB 1399 CCACCCCGTAGAGGAGAGGACATCCACACACTCCGTGAGAGTCCAGAGGCGCCAGCA 1458
 QY 161 G1YVPROG1YGL1YVMECTG1YVARG1YGL1YVARG1YGL1YVARG1YGL1YVARG1Y 180
 DB 1459 GGTCTCTGGGAGACCATGAGTGGATGGAGGCGGTGAGAGATGAAGAGGCTTCCCT 1518
 QY 181 PROARG1YVPROARG1YVSE1YVARG1YVAP1YVSE1YVGL1YV1A1N1V1A1G1N1H1E1A1R1G1 200
 DB 1519 CCAGAGAGACCCCGGGGTCTCCGAGGAAACCCCTGAGAGAGAAACGTCACACCGA 1578
 QY 201 A1AG1YAP1YV1YVARG1YV1YVAP1YV1YVARG1YV1YVARG1YV1YVARG1YV1YVARG1Y 220
 DB 1579 GCTGAGAC1YVARG1YV1YVARG1YV1YVARG1YV1YVARG1YV1YVARG1YV1YVARG1Y 1638
 QY 221 G1UCYV1A1N1YV1YVARG1YV1YVARG1YV1YVARG1YV1YVARG1YV1YVARG1Y 240
 DB 1639 GAGTGAACCAAGTGAAGGCGCCCAAGAGGCTCTCCCGCCACCCCTTCCGCC 1698
 QY 241 PROG1YV1YVARG1YV1YVARG1YV1YVARG1YV1YVARG1YV1YVARG1YV1YVARG1Y 260
 DB 1699 CCGGAGTGAATCTGAGAGAGG1YV1YVARG1YV1YVARG1YV1YVARG1YV1YVARG1Y 1758
 QY 261 ME1YV1YVARG1YV1YVARG1YV1YVARG1YV1YVARG1YV1YVARG1YV1YVARG1Y 280
 DB 1759 ATGATGTGTGTGTCTCCGGTGAATTTCAAGAGTGGCGGTGTGAGACAGAGTGGC 1818
 DB 1819 TTCGTGTGTGTGTCTCCGGTGAATTTCAAGAGTGGCGGTGTGAGACAGAGTGGCCT 1878
 QY 301 G1YGLV1YVPROG1YV1YVARG1YV1YVARG1YV1YVARG1YV1YVARG1YV1YVARG1Y 320
 DB 1879 GGGGGGGCCCCCTGAGCTTTGATGAACAGATGGAGAGAAAGAGAGAGAGAGAGAGAGAGAG 1938
 QY 321 PROG1YV1YVMECTAS1YV1YVARG1YV1YVARG1YV1YVARG1YV1YVARG1YV1YVARG1Y 338

DB 1939 CCGTGAATAATGATTAAGCGAGACGCTCAGAGAGCGAGATCGCCCTAC 1992
 RESULT 4
 ADQ86032
 ID ADQ86032 standard; cDNA; 2372 BP.
 XX
 AC ADQ86032;
 DT 07-OCT-2004 (first entry)
 DE Human tumour-associated antigenic target (TAT) cDNA sequence #2904.
 XX human; tumour-associated antigenic target; TAT; cytostatic; gene therapy;
 XX cancer; cell proliferative disorder; gene; ss.
 OS Homo sapiens.
 PN W02004060270-A2.
 XX
 PD 22-JUL-2004.
 XX
 PF 15-OCT-2003; 2003WO-US029126.
 XX
 PR 18-OCT-2002; 2002US-0418988B.
 PA (GENTH) GENENTECH INC.
 PA (WUTD/) WU T D.
 PA (ZHOU/) ZHOU Y.
 PI WU TD, Zhou Y;
 XX
 DR WPI; 2004-534300/51.
 XX
 PT New nucleic acid molecule and encoded polypeptide, for diagnosing,
 PT preventing or treating cell proliferative disorders such as cancer.
 PS
 XX
 PS Claim 1; SEQ ID NO 2904; 5504bp; English.
 XX
 CC The present invention describes an isolated tumour-associated antigenic
 CC target (TAT) nucleic acid comprising: (a) any of 4622 nucleotide
 CC sequences (see SEQ ID NO:1 to 4622); (b) the full-length coding region of
 CC (a); (c) the complement of (a) or (b); (d) a sequence that has 80%
 CC sequence identity to (a)-(c); or (e) a sequence that hybridizes to (a)-
 CC (c). Also described: (1) an expression vector comprising the above
 CC nucleic acid; (2) a host cell comprising the above expression vector; (3)
 CC a process for producing a polypeptide; (4) an isolated polypeptide
 CC comprising: (a) an amino acid sequence encoded by any of the above
 CC nucleotide sequences; (b) an amino acid sequence encoded by the full-
 CC length coding region of the above nucleotide sequences; or (c) a sequence
 CC having at least 80% identical to (a) or (b); (5) a chimeric polypeptide
 CC comprising the above polypeptide fused to a heterologous polypeptide; (6)
 CC an isolated antibody that binds to the above polypeptide; (7) a process
 CC for producing the antibody; (8) an isolated oligopeptide that binds to
 CC the above polypeptide; (9) a tumour-associated antigenic target (TAT)
 CC binding organic molecule that binds to the above polypeptide; (10) a
 CC composition of matter comprising the above (chimeric) polypeptide;
 CC antibody, oligopeptide or TAT binding organic molecule, in combination
 CC with a carrier; (11) an article of manufacture comprising: (12) methods and
 CC the composition of matter contained within the container; (12) methods of
 CC inhibiting the growth of a cell that expresses the above protein, where
 CC the growth of the cell is at least in part dependent upon a growth
 CC potentiating effect of the above protein; (13) a method of
 CC therapeutically treating a mammal having a cancerous tumour comprising
 CC cells that express the above protein; (14) a method of determining the
 CC presence of a protein in a sample suspected of containing the protein
 CC described above; (15) methods of diagnosing the presence of a tumour in a
 CC mammal; (16) a method for treating or preventing a cell proliferative
 CC disorder associated with increased expression or activity of the above
 CC protein; and (17) a method of binding an antibody, oligopeptide or
 CC organic molecule to a cell that expresses the protein described above.
 CC The TAT sequences have cytostatic activities, and can be used in gene
 CC therapy. The composition and methods are useful for diagnosing,
 CC preventing or treating cancer. The composition is also used for preparing

Alignment Scores:

Prod. No.: 2,14e-60 Length: 2388
 Score: 1922.00 Matches: 338
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 5 Gaps: 0

US-10-791-017a-2_copy_319_656 (1-338) x AAs70647 (1-2388)

QY 1 G1YGLYARGLYGLYMETGLYSERIALAGLYGLUARGLYGLYVHEANLYPARGLYGLY 20
 DB 1015 GGAGGACGCGGTGAATGGGACGCTGGAGACGAGGTGCTTCATTAAGCCCTGGTGA 1074
 QY 21 PRCMETASPG1UGLYPARGLEUAPLEUGLYPROPROVALIAPPAPAPG1UAPPSER 40
 DB 1075 CCCATGGATAGAGACCAAGATCTTGATCTAGGCCCTCTTAATCCAGATGAAGACTCT 1134
 QY 41 ASPASNSERLAIETRYVALG1NGLYLEUASNPASERVALTHLEUAPAPLEUALA 60
 DB 1135 GACAAACAGTCAATTATATGACAAAGATTAATGACAGTGTGACTCTAGATGATCTGGCA 1194
 QY 61 ASPPHEHELPG1NCGYGLYVALLYMEETASNP1YARGTNRGLYGLINPROMETILE 80
 DB 1195 GACTTCTTTAAGCAGTGTGGGGTGTTAAGATGAACAAGAACTGGCAACCCATGATC 1254
 QY 81 H1E1LETRYLEUAPLYG1UHTNRGLYLYPARGLYVAPAPATHVALSERTRYGLU 100
 DB 1255 CACATCTACTGACCAAGAAAGCAAGAAAGCCCAAGACCATGCCACAGTCTCATGA 1314
 QY 101 ASPPROBOTHRALALYSALALALVALGLUTRPHENAPG1LYVAPPHENGLYSER 120
 DB 1315 GACCCACCCGACGCAAGAGGCTGCCGTGGAAATGGTTTATGGGAAAGATTTTCAAGGAGC 1374
 QY 121 LYSLEULYVALSERLEUALAARGLYLYPARGLEUAPSERVALTHLEUAPAPLEUALA 140
 DB 1375 AAACCTTAAGTCTCCCTTCTCGGAAGAACCTTCATGAAGATGCGGGGTGGTCTG 1434
 QY 141 PROPARAG1UGLYARGGLYMETPROPROLEUARG1VGLYVPROGLYVGLYPROGLY 160
 DB 1435 CCACCCCGTGAAGGACAGGAGCATGCAACCACTCGTGAAGGTCCAGAGGACCCCAAGA 1494
 QY 161 G1YPROGLYVGLYPROMETGLYARGMETGLYVARGLYVGLYVAPARGGLYVGLYPHEPRO 180
 DB 1495 GGTCTGGGGGACCCAGTGGGTCCGATGGAGGCCGTGGAGAGATGAGAGAGCCTTCCCT 1554
 QY 181 PROARG1YPRARG1YSEIARG1YASNPSEIARG1YGLYVAPNALGINH1ARG 200
 DB 1555 CCAAGAGGACCCCGGGGTTCGCCAGGGAGAACCCCTCTGAGAGAGAAAGCTCCAGACCGA 1614
 QY 201 ALAG1YASPTTRG1NCGYSPROANPARG1VGLYVAPNALGINH1ARG 220
 DB 1615 GCTGGAGACTGGAGGTGTCCCAATCCGGGTGTGGAAACAGAACTTCGCTGGAGAGCA 1674
 QY 221 GLUCYASNPGLINCYSALAPROLYPARG1VGLYVHEANLYPARG1VGLYVHEANLYPARG1VGLYV 240
 DB 1675 GAGTGCAACAGGTGTAAGGCCCAAGACCTGGAAGGCTTCTCCGCCACCTTTCCGCCCC 1734
 QY 241 PRCGLYVGLYVAPARG1YARGGLYVGLYVARGGLYVGLYVARGGLYVGLYVARGGLYV 260
 DB 1735 CCGGTGTGTATGCTGTCAGAGGTGGTCCCTGTCGTCATGCGGGAGAGAAAGGTGGCTC 1794
 QY 261 METASPARGLYVGLYVARGGLYVGLYVARGGLYVGLYVARGGLYVGLYVARGGLYV 280
 DB 1795 ATGATATGTGTGTCCCGGTGGAATTTCAAGAGTGGCCGTGTGAGACAGAGGTGGC 1854
 QY 281 PHEARG1YGLYVARGGLYVMEETASPARGLYVGLYVHEGLYVGLYVARGGLYVGLYVARGGLYV 300
 DB 1855 TTTCGTGTGTGGCCGGGACATGACCAAGGTGGCTTTGTGTGAGAGAAACAGGTGGCTT 1914
 QY 301 G1YGLYVARGGLYVARGGLYVARGGLYVARGGLYVARGGLYVARGGLYVARGGLYV 320
 DB 1915 GGGGGGGCCCTTGAGCTTTGATGAGACAGATGGAGAGAAAGAGAGAGACGTGGAGGA 1974

QY 321 PROGLYLYMEETASPARGLYVGLYVARGGLYVGLYVARGGLYVGLYVARGGLYV 338
 DB 1975 CTTGGAATAATGATTAAGCCAGACCTCTCAGAGCGCAGAGATCGCCCTTAC 2028

RESULT 6
 ABK84628
 ID ABK84628 standard; cDNA; 2390 BP.
 AC ABK84628;
 XX
 DT 14-AUG-2002 (first entry)
 XX
 DE Human cDNA differentially expressed in granulocytic cells #1199.
 KW Human; se; granulocytic cell; DNA chip; bacterial infection;
 KW viral infection; parasitic infection; protozoal infection;
 KW fungal infection; sterile inflammatory disease; psoriasis;
 KW rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;
 KW cardiac reperfusion injury; renal reperfusion injury; ARDS;
 KW adult respiratory distress syndrome; inflammatory bowel disease;
 KW Crohn's disease; ulcerative colitis; periodontal disease;
 KW granulocyte activation; chronic inflammation; allergy.
 OS Homo sapiens.
 XX
 PN WO200228999-A2.
 XX
 PD 11-APR-2002.
 XX
 PF 03-OCT-2001; 2001WO-US030821.
 XX
 PR 03-OCT-2000; 2000US-0237189P.
 XX
 PA (GENE-) GENE LOGIC INC.
 XX
 PI Beazer-Barclay Y, Weissman SM, Yamaga S, Vockley J;
 XX
 DR WPI; 2002-435328/46.
 XX
 PT Detecting granulocyte activation by detecting differential expression of
 PT genes associated with granulocyte activation, which serves as diagnostic
 PT markers that is useful for monitoring disease states and drug toxicity.
 XX
 PS Claim 1; SEQ ID NO 1199; 114pp; English.

The invention relates to detecting (M1) granulocyte (GC) activation (GCA), by detecting the level of expression of gene(s) (Gs) identified by GCA chip analysis as given in the specification, and comparing the expression level to an expression level in an unactivated GC, where differential expression of Gs is indicative of GCA. Also included are modulating (M2) Gs by contacting GC with an agent that alters the expression of at least one gene in Gs; (2) screening (M3) for an agent capable of modulating GCA or an inflammation (especially chronic) in a tissue, an allergic response in a subject, exposure of a subject to a pathogen or sterile inflammatory disease using the gene expression profile; (3) detecting (M4) an inflammation (especially chronic) in a tissue, an allergic response in a subject, exposure of a subject to a pathogen or sterile inflammatory disease, by detecting the level of expression in a sample of the tissue of gene(s) from Gs, where the level of expression of the gene is indicative of inflammation; (4) treating (M5) an inflammation (especially chronic) or in a tissue, an allergic response in a subject, exposure of a subject to a pathogen or sterile inflammatory disease, by contacting a tissue having inflammation with an agent that modulates the expression of gene(s) from Gs in the tissue. M1 is useful for detecting GCA; M2 is useful for modulating GCA; M3 is useful for screening an agent capable of modulating GCA preferably in an inflammation in a tissue; M4 is useful for detecting an inflammation (especially chronic) in a tissue, an allergic response in a subject, exposure of a subject to a pathogen or sterile inflammatory disease (e.g. psoriasis, rheumatoid arthritis, glomerulonephritis, asthma, thrombosis, cardiac reperfusion injury, renal reperfusion injury, ARDS, adult respiratory distress syndrome, inflammatory bowel disease, Crohn's

CC prostate cancer, ABK64106-ABK64860 represent human benign prostatic
CC hyperplasia gene sequences of the invention
XX
SQ Sequence 2390 BP, 645 A, 589 C, 668 G, 488 T, 0 U, 0 Other;

Alignment Scores:

Pred. No.:	2,14e-60	Length:	2390
Score:	1922.00	Matches:	338
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	6	Gaps:	0

US-10-791-017a-2_COPY_319_656 (1-338) x ABK64822 (1-2390)

QY 1 GYGLYARGLYGLYMETGLYSERIALAGLYGLUARGLYGLYPHEANLYSPROGLYGLY 20
DB 998 GGAGGACGCGGTGAATGGCGACGCTGGAGAGCGTGGCTTCAATAAGCCCTGGTGA 1057
QY 21 PROMETASPGIUGLYPROAEPLEUAPLEUGLYPROBVALAPPROAEPGLUAPSER 40
DB 1058 CCCATGATAGACACCAATCTTGATCTAGCCCTCTGTAGATCCAGATGAAGACTCT 1117
QY 41 ASPASERIALIETRYVALINGLYLEUASAPSERVALTHLEUAPAPLEUALA 60
DB 1118 GACACAGTCGCAATTATATGACAGATTAATGACAGTGTGACTCTAGATGATCTGGCA 1177
QY 61 APPHEPHELYGLINCYGLYVALLYLYMECHANLYSARGTHRYGLINPROMETILE 80
DB 1178 GACTCTTTAAGCAGTGTGGGTGTGTTAAGATGAAACAGAGAACTGGGCAACCCATGATC 1237
QY 81 HIGILETRYLEUAPLYSGIUGLYTHRYGLYSPROLYGLYAPALATHVALSERTYRGU 100
DB 1238 CACATCACTCGACCAAGAAACAGAAAGCCCAAGCCATGACACAGTCTATGAA 1297
QY 101 APPPROTHRALALYSALALAVAGLYUTRPHASPGLYLYSAPPHEGLINGLYSER 120
DB 1298 GACCCACCCACCTGCGAAGGCTGCGGTGGAAATGGTTGATGGAAAGATTTTCAAGGAGC 1357
QY 121 LYLEULYVALSERLEUALATRGLYLYSPROBMECHANSEMERVALGLYGLYLEU 140
DB 1358 AAACCTTAAGTCTCCCTGCTCGGAGAAACCTTCAATGAACGATAGCCGGGTGTCTG 1417
QY 141 PROBARGIUGLYARGLYMETPROBPROBLEUARGLYGLYPROGLYGLYPROGLY 160
DB 1418 CACCCCGTAGAGGACAGAGGATGCAACCACTCCGTGAGGTCCAGAGGCGCCAGGA 1477
QY 161 GLYPROGLYGLYPROMETGLYARGMETGLYGLYARGLYGLYAPARGLYGLYPHEPRO 180
DB 1478 GGTCTGGGGGACCCCATGGGTCCGATGGAGGCGCTGAGAGATGAGAGGCTTCCCT 1537
QY 181 PROARGLYPROARGLYSERARGLYASMPROSERGLYGLYVAENVALGINHISARG 200
DB 1538 CCAAGAGAGACCCCGGGGTTCGCCGAGAACCCCTCTGAGAGAGGAAACGTCACGACCGA 1597
QY 201 ALGLYASPTRGINCYSPROANPROGLYCYGLYVANGLYANPHEALATPARGTTH 220
DB 1598 GCTGGACACTGGCAGTGTCCCAATCCCGAGTTGTAAGAACAGAACTTGGCTGGAGACA 1657
QY 221 GLUCYANGLINCYGLYBALAPROLYPROGLYGLYGLYGLYGLYGLYGLYGLYGLY 240
DB 1658 GAGTGCAACCACTGTAAGGCGCCCAAGCCCTGAGAGGCTTCCCTCCGCAACCTTCCGCC 1717
QY 241 PROGLYGLYAPARGLYGLYARGLYGLYPROGLYGLYMETARGLYGLYGLYGLYGLY 260
DB 1718 CCGGTGTGTATCTGTGACAGAGGTGGCTGTGTGTGATGCGGGAGAGAGAGGTGGCTC 1777
QY 261 METASPARGLYGLYPROGLYGLYMETPHARGLYGLYGLYARGLYGLYAPARGLYGLY 280
DB 1778 ATGATGTGTGTGTCCCGGTGGATGTTCAAGAGTGTCCGTGTGAGACAGGTGGC 1837
QY 281 PHEARGLYGLYARGLYMETASPARGLYGLYGLYGLYGLYGLYGLYGLYGLYGLYGLY 300

DB 1838 TTCCTGTGTGCGGGGACATGACCGAGGTGCTTTGTGTGAGAGAAAGACAGGTGCCCT 1897
QY 301 GYGLYPROBPROGLYPROLEUMETGLIUGLYMETGLYGLYARGARGLYGLYARGGLYGLY 320
DB 1898 GGGGGGCCCCCTGGACCTTTGATGAAACAGATGGGAGAGAGAGAGAGAGAGAGAGAG 1957
QY 321 PROGLYLYMECASPLYSGIUGLYHISARGINGLYLARGARGSPARPROTYR 338
DB 1958 CTTGAAAAATGATTAAGCGACGACCTTCAGAGCGCAGAGATCGGCCCTTAC 2011

RESULT 8

ABN97274 standard; DNA; 2390 BP.

ABN97274;

13-AUG-2002 (first entry)

Gene #3772 used to diagnose liver cancer.

KW Gene; liver cancer; de; hepatocellular carcinoma; hepatotropic;
KW metastatic liver tumor; cytostatic; expression profile; disease state;
KW disease progression; drug toxicity; drug efficacy; drug metabolism.

OS Homo sapiens.

PN W0200229103-A2.

PD 11-APR-2002.

PF 02-OCT-2001; 2001WO-US030589.

PR 02-OCT-2000; 2000US-0237054P.

(GENE-) GENE LOGIC INC.

PI Horne D, Alvares C, Peres-Da-Silva S, Vockley JG;

DR WPI; 2002-426119/45.

PT Diagnosing and detecting the progression of liver cancer, hepatocellular
PT carcinoma or metastatic liver tumor in a patient, involves detecting the
PT level of expression of two or more genes in a liver tissue sample.

PS Claim 1; SEQ ID NO 3772; 298bp; English.

XX The invention relates to a novel method for diagnosing and detecting the
XX progression of liver cancer, hepatocellular carcinoma or metastatic liver
XX tumor in a patient, and differentiating metastatic liver cancer from
XX hepatocellular carcinoma in a patient, involving detecting the level of
XX expression of two or more genes represented in ABN93503-ABN97455 in a
XX tissue sample. The method of the invention has hepatotropic, and
XX cytostatic activity. The method is useful for diagnosing and detecting
XX the progression of liver cancer, hepatocellular carcinoma and metastatic
XX liver carcinoma in a patient. The method is useful for identifying
XX expression profiles which serve as useful diagnostic markers as well as
XX markers that can be used to monitor disease states, disease progression,
XX drug toxicity, drug efficacy and drug metabolism. Note: The sequence data
XX for this patent did not form part of the printed specification, but was
XX obtained in electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences

SQ Sequence 2390 BP, 645 A, 589 C, 668 G, 488 T, 0 U, 0 Other;

Alignment Scores:

Pred. No.:	2,14e-60	Length:	2390
Score:	1922.00	Matches:	338
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	6	Gaps:	0

US-10-791-017a-2_COPY_319_656 (1-338) x ABN97274 (1-2390)

QY 1 G|YGLYARGLYGLYMETGLYSEALAGLYGLUARGLYGLYVPHENALYPROGLYGLY 20
 DB 998 GGAGAGCGCGGTGGAAATGGGCAAGCGCTGGAGAGCGAGTGGCTTCAATAAGCCCTGGTGA 1057
 QY 21 PrometAspGluGlyProAspLeuAspLeuGlyProProValAspProAspGluAspSer 40
 DB 1058 CCCATGATGAAGAGCAAGATCTTGATCTAGGCCCTCTGTAGATCCAGATGAAGATCTT 1117
 QY 41 AspAsnSerAlaIleTyrValGlnGlyLeuAsnAspSerValThrLeuAspAspLeuAla 60
 DB 1118 GACACAGCTGCAATTTATGTACAAGATTAATGACAGTGTGACTAGATGATCTGCA 1177
 QY 61 AspPheMetGlyGlnGlyValValLysMetAsnLysArgThrGlyGlnProMetIle 80
 DB 1178 GACTCTTTAAGCAGTGTGGGTGTTTAAATGAAGAACATGGGCAACCCATATC 1237
 QY 81 HisIleTyrLeuAspLysGluThrGlyLysProLysGlyAspAlaThrValSerTyrGlu 100
 DB 1238 CACATCTACCTGACAAAGAAACAGAAAGCCAAAGCCATGCCACAGTGTCTTATGA 1297
 QY 101 AspProProThrAlaLysAlaAlaValGluTyrPheAspGlyLysAspPheGlnGlySer 120
 DB 1298 GACCAACCCACTGCGCAAGCGCTGCCGTGAATGGTTTATGGGAAAGATTTTCAAGGAGC 1357
 QY 121 LysLeuLysValSerLeuAlaArgLysLysProProMetAsnSerMetArgGlyGlyLeu 140
 DB 1358 AAACCTTAAGTCTCCCTGCTGGAGAACCTTCAATGAACAGTAAAGCGGGGTGGTCTG 1417
 QY 141 ProProArgGluGlyArgGlyLysMetProProProLeuArgGlyGlyProGlyGlyProGly 160
 DB 1418 CCAACCCGTGAGGCGAAGGCAAGGCAAGCCACACTCCGTGAGAGTCCAGAGAGGCCCA 1477
 QY 161 GlyProGlyGlyProMetGlyArgMetGlyGlyValArgGlyGlyAspArgGlyGlyPhePro 180
 DB 1478 GGTCTGGGGGAGCCCATGGGTGGATGGAGAGCGGTGAGAGATAGAGAGGGCTTCCCT 1537
 QY 181 ProArgGlyProArgGlySerArgGlyAsnProSerGlyGlyValAsnValGlnHisArg 200
 DB 1538 CCAAGAGAGCCCGGGGTTCCTCCAGAGGAAACCTCTGAGAGAGAAACGCTCAGCACCA 1597
 QY 201 AlaGlyAspTyrGlnCysProAsnProGlyCysGlyValAsnGlnAsnPheAlaTyrPargThr 220
 DB 1598 GCTGGAGACTGGCAAGTCTCCCAATCCGGGTTGTGGAAACAGAACTTCCCTGGAGACA 1657
 QY 221 GlnCysAsnGlnCysLysAlaProLysProGlyGlyPheLeuProProPheProPro 240
 DB 1658 GAGTGAACACAGTGAAGGCCCAAGCCCTGAGAGCTTCCCGCACCTTTCCGCC 1717
 QY 241 ProGlyGlyAspArgGlyArgGlyGlyProGlyGlyMetArgGlyGlyValGlyLeu 260
 DB 1718 CCGGAGTGTATGTTGCAAGAGTGGCCCTGGTGGCAATGGGAGAAAGAGTGGCTC 1777
 QY 261 MetAspArgGlyGlyLysProGlyGlyMetPheArgGlyGlyValArgGlyGlyAspArgGlyGly 280
 DB 1778 ATGATGCTGTGTCTCCGATGATGTTCAGAGGTGGCCGTGTGAGACAGAGGTGGC 1837
 QY 281 PheArgGlyGlyValArgGlyMetAspArgGlyGlyPheGlyGlyGlyValArgArgGlyGlyPro 300
 DB 1838 TTCGTGTGTGGCGGGGCATGACCGAGGTGGCTTTGTGTGAGAGAGACAGAGGTGGCCT 1897
 QY 301 GlyGlyProProGlyProLeuMetGlnGlnMetGlyGlyValArgArgGlyGlyValArgGlyGly 320
 DB 1898 GGGGGGCGCCCTGAGACTTTGATGAGAACATAGGAGAAAGAGAGAGAGAGAGAGAGAG 1957
 QY 321 ProGlyLysMetAspLysGlyGlnHisArgGlnLysArgAspArgProTyr 338
 DB 1958 CTTGAAATTAATGAATTAAGGCGAGCACGTCAGAGCGCAGAGTGGCCCTAC 2011
 RESULT 9
 ADS16277
 ID ADS16277 standard; cDNA; 2390 BP.
 XX

AC ADS16277;
 DX 18-NOV-2004 (first entry)
 DX
 DE Human cDNA encoding the Ewing sarcoma protein SegID 1.
 XX
 KW human; sei gene; Ewing sarcoma; EWS; prostatic cancer; alopecia; acne;
 KW hypogonadism; androgen-resistance syndrome; testicular feminisation.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT CDS 44..2014
 FT /*tag= a
 FT /product= "Ewing sarcoma protein"
 XX
 PN EP145190-A1.
 XX
 PD 08-SEP-2004.
 XX
 PF 16-FEB-2004; 2004EP-00003422.
 XX
 PR 04-MAR-2003; 2003DE-01009280.
 PR 25-APR-2003; 2003US-0465692P.
 XX
 PA (SCHD) SCHERING AG.
 XX
 XX Obendorf M, Wolf S;
 PI
 DR WPI; 2004-627861/61.
 DR P-PSDB; ADS16278.
 DX
 PT Determining the hormonal effects of substances, used to identify
 PT pharmaceuticals, e.g. for treatment of androgen receptor dysfunction,
 PT from modulating interaction between nuclear receptors and Ewing sarcoma
 PT protein.
 XX
 PS Claim 7, SEQ ID NO 1, 30pp; German.
 XX
 CC This invention relates to a novel modulators that alter the interaction
 CC between the Ewing sarcoma protein (EWS) and its nuclear receptor, as well
 CC as the screening method thereof. Specifically, it refers to determining
 CC and identifying a hormonal effect brought about by test compounds that
 CC modulate either the binding of EWS to the nuclear receptor or the ligand-
 CC induced activity of this receptor. The present invention describes the
 CC nuclear receptors as including oestrogen, progesterone, thyroid hormone,
 CC Vitamin D, and retinoic acid receptors, most preferably they are androgen
 CC receptors. Accordingly, these modulators may be used in the development
 CC of pharmaceutical compositions that can diagnose and be used to treat
 CC diseases associated with receptor dysfunction such as prostatic cancer,
 CC alopecia, acne, hypogonadism and androgen-resistance syndrome e.g.
 CC testicular feminisation. This method provides reliable, sensitive,
 CC simple, inexpensive and rapid assessment of the hormonal effects of these
 CC test compounds. This polynucleotide sequence is the cDNA encoding the
 CC human Ewing sarcoma protein of the invention.
 XX
 SQ Sequence 2390 BP; 646 A; 589 C; 668 G; 487 T; 0 U; 0 Other;
 XX
 Alignment Scores:
 Pred. No.: 2,14e-60 Length: 2390
 Score: 1922.00 Matches: 338
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 13 Gaps: 0
 US-10-791-017A-2_COPY_319_656 (1-338) X ADS16277 (1-2390)
 QY 1 G|YGLYARGLYGLYMETGLYSEALAGLYGLUARGLYGLYVPHENALYPROGLYGLY 20
 DB 998 GGAGAGCGCGGTGGAAATGGGCAAGCGCTGGAGAGCGAGTGGCTTCAATAAGCCCTGGTGA 1057
 QY 21 PrometAspGluGlyProAspLeuAspLeuGlyProProValAspProAspGluAspSer 40

```

Db      1058  CCATGATGATAGAGACCAAGATCTTGTATCTGAGCCCACTGATGATCCAGATGAAGACTCT 1117
Qy      41    AspAsnSerAlaIleTyrValGlnGlyLeuAsnAspSerValThrLeuAspLeuAla 60
Db      1118  GACAAAGCTGCAATTTATGTACAAAGATTAATGACAGTGTGACTGATGATGATCTGGCA 1177
Qy      61    AspPhePheLeuGlnGlyValValValMetAsnLysArgThrGlyGlnProMetIle 80
Db      1178  GACTTCTTTAAGAGTGTGGGTTGTTAAGATGAACAGAACTGGGCAACCATGATC 1237
Qy      81    HisIleTyrLeuAspLysGlnThrGlyLysProLysGlyAspAlaThrValSerTyrGlu 100
Db      1238  CACATCTACCTGACAAAGAAACAGAAACCCAAAGCCATGCCACAGTCTCTATGAA 1297
Qy      101    AspProProThrAlaLysValAlaValGluTyrPheAspGlyLysAspPheGlnGlySer 120
Db      1298  GACCCACCCCACTGCGCAAGGCTGCGGTGAATGGTTGATGGGAAAGATTTCAGAGGAGC 1357
Qy      121    LysLeuLysValSerLeuAlaArgLysValProPheMetAsnSerMetArgGlyLysLeu 140
Db      1358  AAACCTTAAGTCTCCCTTGTCTCGAAGAGAGCTTCAATGAACAGTATGCGGGGTGCTG 1417
Qy      141    ProProArgGlnGlyLysArgLysMetProProProLeuArgGlyLysProGlyLysPro 160
Db      1418  CCACCCCGTGAAGGCGAGAGGACATGCCACCACTCCGTGAGGTCCAGAGAGGCCAGGA 1477
Qy      161    GlyProGlyLysLysProMetGlyLysArgMetGlyLysValArgGlyLysAspArgGly 180
Db      1478  GGTCTCTGGGAGACCATGGGTCCGATGGAGGCGGTGAGAGAGATGAGAGAGGCTTCCCT 1537
Qy      181    ProArgGlyLysProArgGlyLysSerArgGlyLysProSerGlyLysGlyAsnValGlnHis 200
Db      1538  CCAGAGAGAGCCCGGGGTTCCCGAGGGAACCCCTCTGAGAGAGAAAGTCCAGACCGGA 1597
Qy      201    AlaGlyAspTyrPyrGlnCysProAsnProGlyLysGlyLysValGlnAsnPheAlaTyr 220
Db      1598  GCTGGAGACTGGCAGTGTCCCAATCCGGGTTGTGGAACCAAGAACTTCGCTGGAGAA 1657
Qy      221    GluCysAsnGlnCysValLysAlaProLysProGlnGlyLysLeuProProPheProPro 240
Db      1658  GAGTGAACCAAGTGTAAAGGCCCCCAAGAGGCTTCTCCGCAACCCCTTCCGCCCC 1717
Qy      241    ProGlyLysLysAspArgGlyLysArgGlyLysProGlyLysMetArgGlyLysArgGly 260
Db      1718  CCGGCTGTGTATCTGGCAGAGGAGTGGCCCTGGTGGCAATGCGGGAGGAAGGTTGGCTC 1777
Qy      261    MetAspArgGlyLysProGlyLysMetPheArgGlyLysValArgGlyLysAspArgGly 280
Db      1778  ATGATGTGTGTGTGTCCCGGTGGAATTTCAAGAGTGGCGGTGTGAGACAGAGGTGGC 1837
Qy      281    PheArgGlyLysValArgGlyLysMetAspArgGlyLysPheGlyLysValArgGlyLys 300
Db      1838  TTCGCTGTGTGGCCGGGCGCATGACCGGAGTGGCTTTGTGTGAGAAACAGAGGTGGCCCT 1897
Qy      301    GlyLysLysProProGlyLysProLeuMetGlnGlyLysValArgGlyLysValArgGly 320
Db      1898  GGGGGGGCCCTTGAGCTTTGATGAAACAGTGGAGAGAGAGAGAGAGAGAGCGTGGAGGA 1957
Qy      321    ProGlyLysMetAspLysGlyLysLysValArgGlnGlnLysArgAspArgProTyr 338
Db      1958  CTTGGAAAATGATTAAGAGGAGACACCGTCAAGAGCGCAGAGATCCGCCCTTAC 2011

```

RESULT 10

ACN40903 ID ACN40903 standard; cDNA; 2390 BP.

XX AC ACN40903;

XX DT 18-NOV-2004 (first entry)

XX DE Tumour-associated antigenic target (TAT) cDNA DNA269830, SEQ ID NO:5989.

XX

KM Tumour-associated antigenic target; TAT; human; overexpression; cancer;
 KM tumour; diagnosis; cell proliferative disorder; breast cancer;
 KM colorectal cancer; lung cancer; ovarian cancer; liver cancer;
 KM central nervous system cancer; bladder cancer; pancreatic cancer;
 KM cervical cancer; melanoma; leukemia; hybridisation probe;
 KM chromosome identification; chromosome mapping; gene mapping;
 KM gene therapy; cytostatic; gene; ss.

OS Homo sapiens.
 XX
 XX WO2004030615-A2.
 XX
 XX 15-APR-2004.
 XX
 XX 29-SEP-2003; 2003WO-US028547.
 XX
 XX 02-OCT-2002; 2002US-0414971P.
 XX
 XX (GENTH) GENENTECH INC.
 XX
 XX Wu TD, Zhang Z, Zhou Y;
 XX
 XX WPI; 2004-347921/32.
 XX
 XX P-PSDB; ABM82330.
 XX
 XX New tumor-associated antigenic target polypeptides and nucleic acids,
 PT useful in preparing a medicament for treating or detecting a
 PT proliferative disorder, e.g. breast, lung, colorectal, ovarian or
 PT prostate cancer or tumor.
 XX
 XX Claim 1; SEQ ID NO 5989; 7273bp; English.

CC The invention relates to human tumour-associated antigenic target (TAT)
 CC polypeptides, and their related nucleic acids. The TAT polypeptides are
 CC overexpressed in cancer tissues compared to normal tissues, and may thus
 CC serve as effective targets for the diagnosis and treatment of cancer in
 CC mammals. The invention also relates to nucleic acid and polypeptide
 CC sequences at least 80% identical to the TAT nucleic acids and
 CC polypeptides; expression vectors and host cells comprising a TAT nucleic
 CC acid; an antibody specific for a TAT polypeptide; a peptide or organic
 CC molecule which binds to a TAT polypeptide; fusion proteins comprising a
 CC TAT polypeptide; and methods and compositions for the treatment or
 CC diagnosis of cancer in mammals. TAT polypeptides, nucleic acids,
 CC antibodies, antagonists, binding molecules and compositions are useful
 CC for diagnosing or treating a cell proliferative disorder associated with
 CC increased TAT expression, particularly cancers such as breast cancer,
 CC colorectal cancer, lung cancer, ovarian cancer, liver cancer, bladder
 CC cancer, pancreatic cancer, cervical cancer, cancers of the central
 CC nervous system, melanoma and leukemia. TAT nucleic acids may further be
 CC used as hybridisation probes, in chromosome and gene mapping, in
 CC chromosome identification and in gene therapy. The present sequence
 CC represents a TAT nucleic acid of the invention.

XX
 XX Sequence 2390 BP; 645 A; 589 C; 668 G; 488 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	2,14e-60	Length:	2390
Score:	1922.00	Matches:	338
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	13	Gaps:	0

US-10-791-017A-2_COPY_319_656 (1-338) x ACN40903 (1-2390)

```

Qy      1    GlyLysArgGlyLysLysMetGlySerAlaGlyLysValArgGlyLysPheAsnLysProGly 20
Db      998  GAGAGACGGGTGGAATGGGCAAGCGCTGGAGAGCGAGTGGCTTCAATAAGCCGTGGTGA 1057
Qy      21    ProMetAspGlnGlyLysProAspLeuAspLeuGlyProProValAspProAspGlyAsp 40
Db      1058  CCATGATGATGAGACCAAGATCTTGTATCTGAGCCCTCTGTATGATCCAGATGAAGACTCT 1117

```

QY 41 AspAsnSerAlaIleTyrValGlnGlyLeuAsnAspSerValThrLeuAspLeuAla 60
 DB 1118 GACAAAGAGTCATTTATATGACAGATTAATGACGTGTGACTAGATGATCTGGCA 1177
 QY 61 AspPhePheIleGlnCysGlyValValIleMetAsnIleArgThrGlyGlnPrometIle 80
 DB 1178 GACTTCTTTAAGACAGTGTGGGTTGTTAAGATGACAGAGAACTGGGCAACCATGATC 1237
 QY 81 HisIleTyrLeuAspIleGlyIleThrGlyLeuPheProIleGlyAspAlaThrValSerTyrGlu 100
 DB 1238 CACATCTACCTGACAAAGAAACAGGAAACCCAAAGGCATGCCACAGTGTCTATGAA 1297
 QY 101 AspProProThrAlaIleValAlaValGluTyrPheAspGlyIleAspPheGlnGlySer 120
 DB 1298 GACCCACCCCACTCCCAAGGCTGCCGGAATGGTTGATGGAAAGATTTTCAAGGGAGC 1357
 QY 121 LysLeuIleValSerLeuAlaArgLysLysPheProMetAsnSerMetArgGlyIleLeu 140
 DB 1358 AAACCTTAAGTCTCCCTGCTCCGAAAGACCTCCATGAAACATATGCGGGTGGTCTG 1417
 QY 141 ProProArgGluIleValArgIleMetProProProLeuArgGlyIleProGlyIleProGly 160
 DB 1418 CCAACCCGAGGAGGAGGAGCATGCAACCACTCGTGAAGTCCAGAGAGGCCACGGA 1477
 QY 161 GlyProGlyIleValProMetGlyValArgMetGlyIleValArgIleValAspArgGlyIlePhePro 180
 DB 1478 GGTCTCGGGGAGCCATGGGTCCATGGAGGAGCCGTGAGAAATGAGAGAGGCTTCCCT 1537
 QY 181 ProArgGlyProArgGlySerArgGlyAsnProSerGlyIleGlyIleValIleGlnIleAspG 200
 DB 1538 CCAAGAGAGACCCCGGGGTTCCCAAGAGAACCCCTCTGAGAGAGAAAGCTTCACGACCA 1597
 QY 201 AlaGlyAspIlePheIleCysProAsnProGlyIleCysGlyAsnGlnAsnPheAlaIlePargThr 220
 DB 1598 GCTGAGAGATCGGACAGTGTCCCATCCGGGTTGGAAACCAAGAACTTCGCTGAGAGACA 1657
 QY 221 GluCysAsnGlnCysLysAlaIleProIlePheLeuProProPheProPro 240
 DB 1658 GAGTGCACCAAGGTAAAGGCCCAAGGCTTCCCTCCGCAACCCCTTCCGCC 1717
 QY 241 ProGlyIleValAspArgGlyArgGlyIleProGlyIleMetArgGlyIleValArgGlyIleLeu 260
 DB 1718 CCGGCTGGTATCTGGCAGAGGTGGCCCTGTGTGGCATGCGGAGAGAGAGAGGTGGCTTC 1777
 QY 261 MetAspArgGlyIleValProGlyIleMetPheArgGlyIleValArgGlyIleValAspArgGlyIle 280
 DB 1778 ATGATCTGTGTGTCCGCTGGAAATGTTCAAGAGTGGCGGTGTGAGACAGAGGTGGC 1837
 QY 281 PheArgGlyIleValArgIleMetAspArgGlyIlePheGlyIleValArgArgGlyIlePro 300
 DB 1838 TTCCGTGTGGCCGGGCGATGAGACGAGGTGGCTTTGTGTGAGAGAAAGACAGGTGGCCCT 1897
 QY 301 GlyIleProProGlyIleProLeuMetGlnGlnMetGlyIleValArgArgGlyIleValArgGlyIle 320
 DB 1898 GGGGGGGCCCCCTTGATCTTTGATGAAACATGAGAGAAAGAGAGAGAGAGAGAGAGAGAG 1957
 QY 321 ProGlyIleMetAspLysGlyIleHisArgGlnGlnArgArgAspArgProTyr 338
 DB 1958 CCGGAAAAATGATTAAGCGAGCACCGTCAGAGAGGCGAGAGATGGGCCCTTAC 2011
 RESULT 11
 ADR07446
 ID ADR07446 standard; cDNA; 2026 BP.
 AC ADR07446;
 XX
 DT 04-NOV-2004 (first entry)
 XX
 DE Full length human cDNA useful for treating neurological disease Seq 952.
 XX
 KW Gene; 89; human; oligo-capping method; diagnostic marker; gene therapy;
 KM osteoporosis; neurological disease; Alzheimer's disease;
 KM Parkinson's disease; dementia; short memory; cancer;

KW sense or motor function; emotional reaction; fear response; panic;
 KM osteopathic; neuroprotective; nootropic; antiparkinsonian; cyostatic;
 KW tranquilliser.
 XX
 OS Homo sapiens.
 XX
 PN EP1447413-A2.
 XX
 PD 18-AUG-2004.
 XX
 PF 12-FEB-2004; 2004EP-00003145.
 XX
 PR 14-FEB-2003; 2003JP-00102207.
 PR 09-MAY-2003; 2003JP-00131452.
 XX
 PA (REAS-) RES ASSOC BIOTECHNOLOGY.
 XX
 PI Isogai T, Yamamoto J, Nishikawa T, Isono Y, Sugiyama T, Otsuki T;
 PI Makamatsu A, Ishii S, Nagai K, Irie R;
 DR WPI; 2004-583265/57.
 DR P-PSDB; ADR09402.
 XX
 PT New 1995 cDNA, useful for treating osteoporosis, neurological diseases,
 PT Alzheimer's diseases, Parkinson's diseases, dementia and various cancers.
 XX
 PS claim 1; SEQ ID NO 952; 2686pp; English.
 XX
 CC This invention relates to novel, isolated full length human cDNA
 CC molecules and the encoded proteins thereof. Specifically, it refers to
 CC cDNA clones obtained by an oligo-capping method, where none of these
 CC clones are identical to any known human mRNAs. The present invention
 CC describes an immunoassay to identify agonists and antagonists, as well as
 CC antibodies, antisense molecules and siRNAs that can all be used to bind
 CC to and modulate expression of the cDNA molecules. As such, these
 CC molecules are useful for diagnostic markers or therapeutic targets for
 CC the various diseases or morbid states. In particular, they are useful in
 CC gene therapy for treating osteoporosis, neurological disease, Alzheimer's
 CC disease, Parkinson's disease, dementia, short memory and various cancers,
 CC as well as for maintaining equilibrium of sense or motor function, and
 CC for treating emotional reaction, fear response and panic. Accordingly,
 CC they exhibit osteopathic, neuroprotective, nootropic, antiparkinsonian,
 CC cyostatic and tranquilliser activities. This polynucleotide is a full
 CC length human cDNA sequence of the invention. NOTE: This sequence is not
 CC given in the sequence listing of the specification but can be obtained on
 CC CD-ROM from the European Patent Office, Vienna Sub-office.
 XX
 SO Sequence 2026 BP; 540 A; 513 C; 590 G; 383 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 5.53e-60 Length: 2026
 Score: 1909.00 Matches: 337
 Percent Similarity: 99.70% Conservative: 0
 Best Local Similarity: 99.70% Mismatches: 1
 Query Match: 99.32% Indels: 0
 DB: 13 Gaps: 0
 US-10-791-017A-2_COPY_319_656 (1-338) x ADR07446 (1-2026)
 QY 1 GlyIleValArgIleValMetGlySerAlaIleTyrValGlnGlyLeuAsnAspSerValThrLeuAspLeuAla 20
 DB 836 GAGAGAGCGGTGTAATGGCGACGCTGAGAGACGAGGTGGCTTCATTAAGCCTGTGGGA 895
 QY 21 ProMetAspGluIleValProAspLeuAspLeuGlyProProValAspProAspGluAspSer 40
 DB 896 CCAATGATGAAGAGACCAATCTTGAATCTAGGCCCACTGTATATCCAATGAAGACTCT 955
 QY 41 AspAsnSerAlaIleTyrValGlnGlyLeuAsnAspSerValThrLeuAspLeuAla 60
 DB 956 GACAAAGAGTCATTTATATGACAGATTAATGACAGTGTGACTAGATGATCTGGCA 1015
 QY 61 AspPhePheIleGlnCysGlyValValIleMetAsnIleArgThrGlyGlnPrometIle 80

Db		1016	GACTCTCTTAAGCAGTGTGGGATTGTAAATGAACAGAAGAACTGGCCCAACCATGATC	1075		
Oy		81	HtltletYrLleuAaplyvgIuthrclylYspProlyrgslYAspaAlathValSerTyrcLu	100		
Db		1076	CACATCTAACCTGGACAMAGAAAACAGAAAGGCCAAGGCATGCCACAGTGTCTTAGTAA	1135		
Oy		101	AAPProPothrrAlalValAlavalGlutTpPheapsglyLYeaspPheglnglySer	120		
Db		1136	GAACCACCCACTGCGCAAGGCTGCCGTGGAAATGGTTTGATGGGAAAGATTTTCAMGGAGC	1195		
Oy		121	LysleuLyValSerLeuAlarGlyalybProPomeLaanSerMetArsglyVgLyLeu	140		
Db		1196	AAACTTAAGTCTCCCTTGCTCGGAAGAAAGCCTCCAATGAACAGTATCGGGGTGTCTG	1255		
Oy		141	PropoAlargluVArglylYmetProPoleuAtarglyVgLyProglyVgLyProgly	160		
Db		1256	CCACCCCCTGAGGGGAGAGGAGCATGCCACCACTCCGTGGAGGTCCAGAGGGCCCAAGA	1315		
Oy		161	GLYProgiYgLYPrometEgLYArmteGlyvGIYAArggVgLYAspaArggLYgLYPhere	180		
Db		1316	GSCTCTGGGGGACCCATGATGGTGTGCATGGGAGGCGCTGGAGAGATAGAGAGGCTTCCCT	1375		
Oy		181	ProARgLYProARgLYSerARgLYANProsergLYgLYgLYasNValGlnHlsARG	200		
Db		1376	CCAAGAGGACCCCCGGGGTTCGCCAGAGAAACCCCTCTGGAAGAGAGAAAGTCCAGACCGCA	1435		
Oy		201	AlAgLYaBPTrPgInCyBproAnDProGLYCygblYaNgnInanPheAlATrPaRgThr	220		
Db		1436	GCTGGAGACCTGCAGTGTCCCATCTCGGGATTGTGGAAACAGAACTTCGCTGGAGAAACA	1495		
Oy		221	GLUCyBaAngInCYaLYaAlAProLYaBProgiuglylPheLeuProProPheProBo	240		
Db		1496	GAGTGCACCAAGTGTAAAGCCCCCAAAGCCTGAAGGCTTCTCCGCCACCTTTCGGCC	1555		
Oy		241	ProgiYgLYaBpaRGlyARgLYgLYlYProgiYgLYmeArGlylYgLYARgLYgLYLeu	260		
Db		1556	CCGGGTGTGATCGTGCAGAGGTGGCCCTGTGTGCATCGCGGAGAGAAAGAGTGGCTC	1615		
Oy		261	MeaApaRGlygLYlYProgiYgLYlYmetPheaRGlylYgLYARgLYgLYlYAspaRGlylY	280		
Db		1616	ATGATCGTGGTGGTCCCGGTGGAATGTTCAAGAGGTGGCCGGTGGAGAGACAGAGGTGGC	1675		
Oy		281	PheaRGlylYgLYARgLYlYmetAspaRGlylYgLYlYPheglYgLYlYARgARgLYlYPro	300		
Db		1676	TTCGTGGTGGCCCGGGGCAATGACCGAGGTGGCTTGGTGGAGAGAAACGAGGTGGCCCT	1735		
Oy		301	GLYgLYProProgiYProLeuWecGLUInmetGlylYARgARgLYgLYlYARgLYgLYlY	320		
Db		1736	GGGGGGCCCCCTCGAGCTCTTGATGGAAACAGATGGAGAGAAAGAGAGAGAGCTGGAGA	1795		
Oy		321	ProgiLYMeLApLyegLYgLYuhlsARGIngInIwARGARgAspaRgProTYr	338		
Db		1796	CCTGAAAAAATCGATTAAGGCGAGCACCGTCAAGAGCCGACAGAGATCGGCCCTTAC	1849		
RESULT 12						
ID	ABI9383	standard; cDNA; 2188 BP.				
XX	AC	ABI9383;				
DT	07-MAR-2002	(first entry)				
DE	Mouse ischaemic condition related cDNA sequence SEQ ID NO:288.					
KM	Mouse; ischaemia; compressive ischaemia; occlusive ischaemia; vascospastic ischaemia; ischaemic condition; ischaemic disease; ss.					
OS	Mus musculus.					
PN	MO200188188-A2.					
PD	22-NOV-2001.					

Oy		141	PROPFOA	VGGIUGLVARXGILWMECPROFPROPOLEUA	VGGIYGLVPROGI	GLVPROGI	160						
Dd		1428	CCACCTCCTGAAGGCGAGGGGGTATGCCACCACCACTTCGTGAGGCTCTGGTGCCCAAGA				1487						
Oy		161	GlyProGIYgLIYPrometGIYArsMetGIYglYArGIYgLIYAsPARgIYglYphePro				180						
Dd		1488	GGCCCCCTGGAGGAGCCCATGGGTGCATTGGGAGGCCCTGGAGGAGACAGAGGGGGCTTCCCT				1547						
Oy		181	PROARgLIYPROARgLIYSerArgGLYAsnProSerGIYglYlYAsnValGlNHlsArg				200						
Dd		1548	CCAAGAGGGCCCCCGAGGCTCCAGAGAAACCCCTCTGGAGGAGGAGAAATGTCACAGACCGA				1607						
Oy		201	AlAgIYAAPTPIInCYePProasnProGIYCyRgIYAANGlnAsnPhalATTPARGhr				220						
Dd		1608	GCTGGAGACTGGCACTGTCTCCCATCCGGGCTGTGGAAACAGAACTTGCTTGAGAAC				1667						
Oy		221	GIUCYAsnAngINCYelYeAlaProLYSPROGIUGLIYPheLeuProProPhePro				240						
Dd		1668	GAAITCCAAACAAGTGTAAAGGCCCTTAAGCCCAGAGGCTTCTCCGCCACCTTTCCACTT				1722						
Oy		241	PROGIYGIYAAPARgLIYARgLIYglYProGIYglYMeTARgLIYglYARgLIYglYleu				260						
Dd		1728	CCGGGTGTGTATCGTAGGACAGAGGTGGCCCTGTGTGACATGGAGGAGGAAGAGAGACTC				1787						
Oy		261	MeTAPARgLIYgLIYPROGIYglYMeTPhEARgLIYglYAYARgLIYglYASPARgLIYglY				280						
Dd		1788	ATGGAACCGTGTGTGTCTTGAGAGAAATGTTCAAGAGGTGACAGAGGTGGAGACAGAGAGAGC				1847						
Oy		281	PHEARgLIYgLIYARgLIYMeTAsPARgLIYglYPHeGIlYglYlYARARgLIYglYPro				300						
Dd		1848	TTCGAGGTGGCCGTGTGAATGGAACGAGGTGGCTTTGGTGAAGAGAACAGAGTGTCTT				1907						
Oy		301	GIYGIYPROProGIYPROleuMeTglUglinMeTglYglYARgARgLIYglYARgLIYglY				320						
Dd		1908	GCGGGGGCTCCTCGAGCTTTATATGAAACAGATGGAGGAAGAAGGGGAGACGTGAGAGA				1966						
Oy		321	PROGIYlYMeTAsPLYgLIYglUHIsARgINGlnUARgARgPARgPROTr				338						
Dd		1968	CCTGGGAAAAAGAAATAAGCGAGCACCGTCAGAAACGACAGACCGGCCCTAC				2021						
RESULT 13													
ADPS6333													
ID	ADPS6333	standard; cDNA; 1807 BP.											
XX	ADPS6333;												
DT	18-NOV-2004	(first entry)											
DE	Human PRO cDNA sequence SEQ ID NO:2309.												
XX													
KM	human; PRO; immune related disease; inflammatory immune response;												
KM	immune response stimulation; antiallergic; antiinflammatory; antipneumatic;												
KM	antihaemetic; antidiabetic; antiinflammatory; antileptotic;												
KM	antirheumatic; antichryoid; CNS; dermatological; gastrointestinal;												
KM	haemostatic; hepatotropic; immunostimulant; immunosuppressive; muscular;												
KM	nephrotropic; neuroprotective; osteopathic; respiratory; vasotrophic;												
KM	vircicide; gene therapy; gene; ss.												
XX													
OS	Homo sapiens.												
FN	WO2004039956-A2.												
PD	13-MAY-2004.												
XX													
PF	28-OCT-2003; 2003WO-US034381.												
XX													
PR	29-OCT-2002; 2002US-0422472P.												
PA	(GETH) GENENTECH INC.												
XX													
PI	Aggarwal S, Clark H, Gurney AL, Schoenfeld J, Williams PM;												
FI	Wood WL, Wu TD;												
XX													

DR WPI: 2004-376182/35.
DR P-PSDB: ADP56334.
XX
PT New PBO polyamides and polyamides, useful in diagnosing
PT and treating an immune related disease, e.g. systemic lupus
PT erythematosis, rheumatoid arthritis, diabetes mellitus or asthma and in
PT stimulating an immune response.

PS Claim 2; SEQ ID NO 2309; 3009pp; English.

The present invention describes an isolated PRO nucleic acid (1). Also described: (1) a vector comprising (1); (2) a host cell comprising the vector of (1); (3) a process for producing a PRO polypeptides; (4) an isolated PRO polypeptide; (5) a chimeric molecule comprising the polypeptide of (4) fused to a heterologous amino acid sequence; (6) an antibody which specifically binds to a polypeptide of (4); (7) a composition of matter comprising a polypeptide of (4), an agonist or antagonist of the polypeptide or an antibody that binds to the polypeptide in combination with a carrier; (8) an article of manufacture comprising a container, a label on the container and a composition of matter of (7); (9) a method of treating an immune related disease in a mammal; (10) a method for determining the presence of a PRO polypeptide in a sample suspected of having the polypeptide; (11) a method of diagnosing an immune related disease or an inflammatory immune response in a mammal; (12) a method of identifying a compound that inhibits or mimics the activity of or expression of a gene encoding a PRO polypeptide; and (13) a method of stimulating the immune response in a mammal. The PRO sequences have antiallergic, antianemic, antirhectic, antineurathmatic, antidiabetic, antinflammatory, antiparastitic, antineurathmatic, antithyroid, CNS, dermatological, gastrointestinal, haemostatic, hepatotropic, immunostimulant, immunosuppressive, muscular, neurotropic, neuroprotective, osteoporathic, respiratory, vasotropic and vincude activites, and can be used in gene therapy. The nucleic acid (1) and the encoded polypeptides, compositions, kits and methods are useful in diagnosing and treating an immune related disease and in stimulating an immune response. The present sequence represents a human PRO nucleotide sequence from the present invention.

SQ Sequence 1807 BP; 476 A; 497 C; 523 G; 311 T; 0 U; 0 Other;

Alignment Scores:	
Pred. No.:	1,68e-56
Score:	1810.00
Percent Similarity:	100.00%
Best Local Similarity:	100.00%
Query Match:	94.17%
DB:	13
Gaps:	0
Length:	1807
Matches:	318
Conservative:	0
Mismatches:	0
Indels:	0
Gaps:	0

US-10-791-017A-2_COPY_319_656 (1-338) X ADP56333 (1-1807)

Qy	2	ProMetAspGluGlyProAspLeuAspLeuGlyProProValAspProAspGluAspSer	40
Db	851	CCCATGATGAAAGAACCAAGACTTGTAGTACAGCCCTCTGTAGATCCAGATGAAGACTCT	910
Qy	41	AspAsnSerAlaIleTyValGlnGlyLeuAsnAspSerValThrLeuAspAspLeuAla	60
Db	911	GACACACGTCGAATTATTGTACACAGGATTAAATACAGTGTGACTTACATGATCTGGCA	970
Qy	61	AspPhePheLysGlnCysGlyValValLysMetAsnLysArgThrGlyGlnProMetIle	80
Db	971	GACTTCTTTAAGCAGTGTGGGTTGTTAAGATGAACAGAGACTGGGCAACCATGATC	1030
Qy	81	HisIleTyLeuAspLysGluThrGlyLysProLysGlyAspAlaThrValSerTyrGlu	100
Db	1031	CACATCTACCTGGCAACGAAACAGGAAAGCCAAAGCGCATGCCACAGTCTCTTAAGAA	1090
Qy	101	AspProProThrAlaLysAlaAlaValAlaGluTyrPheAspGlyLysAspPheGlnGlySer	120
Db	1091	GACCCACCCACTGCGCAAGGCTGCGCTGGATGTGTTATGGGAAAGATTTTCAAGGAGNC	1150
Qy	121	LysLeuLysValSerLeuAlaArgLysAlaLysProMetAsnSerMetArgGlyGlyLeu	140
Db	1151	AAACTTAAAGTCTCCCTGTGCTGGAAAGAGCTCCATATGAACAGTATGCGGGAGTGCTG	1210

QY	141	ProProAaRgGluGlyVaRgGlyWaeProProProleuAaRgGlyVgLyProGlyVgLyProGly	160		
Db	1211	CCACCCCGTGAAGGCGAGAGCGATGCGACACCACTCCGTGAGGTCCAGAGGCCACGAA	12707		
QY	161	GlyProGlyVgLyProMetGlyVaRgWetGlyGlyVaRgGlyVgLyVaRgGlyVgLyPhePro	180		
Db	1271	GGTCTGGGGAGCCCATGGGTGCGCATGGAGAGCCGTGGAGGAGATGAGAGGCTTCCT	1338		
QY	181	ProARgGlyProARgGlyVseARgGlyVaRnProRseRgLyVgLyVaRnValGlnHiARg	200		
Db	1331	CCAAGAGGACCCCGGGGTTCCTCCGAGGAAACCTCTTGAGGAGGAAAGCTCAGACCGA	1390		
QY	201	AlaGlyAaPTpGlnCyAProAaRnProGlyCyGlyVaRngInAaRnPhaAlaTRPaRgThr	220		
Db	1391	GCTGGAGACTGGCACTGTCTCCCATCTCGGGTTGTGAAACAGAACTTGCTCGAGAAACA	1455		
QY	221	GluCyVaRngInCyELyVaRAlaProLyBProGluGlyPheLeuProProProPheProPro	240		
Db	1451	GAGTGCACCAAGTGTAAAGGCCCAAAAGCCTGAAGGCTTCCTCCGCAACCTTTCGGCC	1510		
QY	241	ProGlyVgLyAaPaRgGlyVaRgGlyVgLyProGlyVgLyMetARgGlyVgLyVaRgGlyVgLyPhe	260		
Db	1511	CCGGGTGTGTATCGTGCGAGAGGTGCCCTGGTGCATGCGGGAGGAAAGAGTGGCTC	15707		
QY	261	MetAaPaRgGlyVgLyProGlyVgLyWaePheARgGlyGlyVaRgGlyVgLyVaRPaRgGly	280		
Db	1571	ATGATCTGTGTGTGTCTCCGCTGGATATGTTCAAGAGGTGGCCGTGTGAGAGACAGAGTGGC	1630		
QY	281	PheARgGlyVgLyVaRgGlyWaeAaPaRgGlyVgLyPheGlyVgLyVgLyVaRgGlyVgLyPro	300		
Db	1631	TTCCCTGTGTGGCCGGGGCATGGAACGAGGTGGCTTTGGTGGAGGAAGCAGAGTGGCCCT	1690		
QY	301	GlyGlyProProProGlyProLeuMetGluGlnMetGlyVgLyVaRgGlyVgLyVaRgGlyVgLy	320		
Db	1691	GGGGGGCCCCCTGGACTTTGTGTGGAACAGATGGAGGAGAGAGAGAGACGTGGAGGA	1750		
QY	321	ProGlyLyMetAaRnLyGlyVgLyHiARgGlnGluARgARgAaRnProTyR	338		
Db	1751	CCTGGAATAATGATTAAGAGGAGCACCGCTCAAGACCAAGATGCGCCCTCAC	1804		
RESULT 14					
ID	AA562623/c	AA562623 standard, cDNA, 2273 BP.			
XX	AA562623;				
DT	14-FEB-2002	(first entry)			
DE	cDNA sequence #410 encoding novel human secreted protein.				
KW	Human secreted protein; hyperproliferative disorder; autoimmune disorder;				
KW	immune deficiency disorder; blood disorder; inflammatory disorder;				
KW	infectious disorder; gene therapy; antimicrobial; hepatotropic;				
KW	immunosuppressive; antirheumatic; ss.				
OS	Homo sapiens.				
PN	WO200177291-A2.				
PD	18-OCT-2001.				
PF	29-MAR-2001; 2001WO-US010485.				
PR	06-APR-2000; 2000US-0195604P.				
PA	(GENY) GENETICS INST INC.				
PI	Wong GG, Clark HF, Fechtel K, Agostino MJ, Howes SH, Resnick RJ,				
DR	Gulikova K, Graham JR;				
XX	WPI; 010900/01.				

[illegible]


```
Db      2778 GGGGGGCCCCCTGGACCTTTGATGAAACAGATGGAGAGAAAGAGAGGACGTGAGCA 2837
Qy      321 ProGlyLysMetAspLys----- 326
Db      2838 CTTGAAAAAATGGATTAAGTCTGTGAAAAAGCAGCTGTGGCCTTAACCGAAGG 2897
Qy      327 -----GlyGluHisArgGluArgArgArgArgArgProTyr 338
Db      2898 CCTCTTTACCTTGCAGAGGCGAGCACGTCAGAGCGCAGAGATCGGCCCTAC 2951
```

Search completed: February 20, 2005, 23:10:21
Job time : 451.771 secs

1115 Page Blank (uspio)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: February 20, 2005, 22:00:21; Search time 149.958 Seconds
(without alignments)
3688.116 Million cell updates/sec

Title: US-10-791-017a-2_COPY_319_656

Perfect score: 1922

Sequence: 1 GAGGCGMSAGRGCGFNKPGS.....GGPGKDKGHRQERDRPY 338

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODE=frame+ p2n.model -DEV=xlp
-O/cgn2_1/USPTO.spool.p/US10791017/runat.17022005.125808.22107/app.query.fasta_1.1358
-DB=Issued Patents.NA -QFMT=fastap -SUFFIX=p2n.rnt -MINMATCH=0.1 -LOOPT=0
-LOOEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blomsu62 -TRANS=human40.cdt
-LIST=45 -DOCLIGN=200 -THR SCORE=epct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10791017@cgn2_1.141@runat.17022005.125808.22107 -NCPU=6 -ICPU=3
-NO MAMP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPELCOPE=100 -LONGLOG
-DEV TIMEOUT=120 -MAIN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database:

Issued Patents.NA.*
1: /cgn2_6/ptodata/1/1na/5A.COMB.seq.*
2: /cgn2_6/ptodata/1/1na/5B.COMB.seq.*
3: /cgn2_6/ptodata/1/1na/6A.COMB.seq.*
4: /cgn2_6/ptodata/1/1na/6B.COMB.seq.*
5: /cgn2_6/ptodata/1/1na/PTCUS.COMB.seq.*
6: /cgn2_6/ptodata/1/1na/backfile1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1922	100.0	2371	2	US-08-343-443B-1
2	1810	94.2	1785	4	US-09-949-016-5043
3	1810	94.2	1785	4	US-09-949-016-5044
4	1632	84.9	6002	4	US-09-949-016-13696
5	1541	80.2	1783	4	US-09-949-016-13696
6	958	49.8	35784	4	US-09-949-016-16785
7	958	49.8	35784	4	US-09-949-016-16785
8	878	45.7	1939	4	US-09-949-016-16786
9	609	31.7	411	4	US-09-621-976-13361
10	405	21.1	601	4	US-09-949-016-176641
11	405	21.1	601	4	US-09-949-016-176641
12	344	17.9	450	3	US-09-370-838-145

13	344	17.9	450	4	US-09-854-133-145	Sequence 145, App
14	289	15.0	1926	3	US-09-249-585A-4	Sequence 4, Appl1
15	289	15.0	1931	2	US-09-130-114-2	Sequence 2, Appl1
16	285	14.8	3147	2	US-08-781-802-7	Sequence 7, Appl1
17	285	14.8	3147	3	US-08-694-078-7	Sequence 7, Appl1
18	285	14.8	3147	3	US-09-058-260-7	Sequence 7, Appl1
19	275	14.3	2082	4	US-09-818-780-67	Sequence 67, Appl
20	272	14.2	4551	4	US-09-949-016-2338	Sequence 238, Ap
21	269.5	14.0	1558	4	US-09-949-016-2635	Sequence 2635, Ap
22	268.5	14.0	4167	4	US-09-169-768-12	Sequence 12, Appl
23	267	13.9	6109	4	US-09-795-061-1	Sequence 1, Appl1
24	266.5	13.9	4568	4	US-09-949-016-2498	Sequence 2498, Ap
25	265	13.8	5387	4	US-09-949-016-265	Sequence 265, App
26	265	13.8	4403765	3	US-09-103-840A-2	Sequence 2, Appl
27	263.5	13.7	1771	2	US-08-533-669A-7	Sequence 7, Appl1
28	263.5	13.7	1771	2	US-08-511-872-1	Sequence 1, Appl1
29	263.5	13.7	1771	3	US-09-183-861-7	Sequence 7, Appl1
30	263.5	13.7	1771	3	US-09-022-765-7	Sequence 7, Appl1
31	263.5	13.7	1771	4	US-09-551-974A-7	Sequence 7, Appl1
32	263.5	13.7	1771	4	US-09-565-501A-7	Sequence 7, Appl1
33	263.5	13.7	1771	4	US-09-639-206A-7	Sequence 7, Appl1
34	263.5	13.7	1771	4	US-09-874-923-7	Sequence 7, Appl1
35	263.5	13.7	1771	4	US-08-798-841-7	Sequence 7, Appl1
36	263.5	13.7	3170	4	US-09-169-768-1	Sequence 1, Appl1
37	263.5	13.7	3171	4	US-09-169-768-15	Sequence 15, Appl
38	263.5	13.7	3171	4	US-09-169-768-19	Sequence 19, Appl
39	263.5	13.7	3349	4	US-09-169-768-13	Sequence 13, Appl
40	263.5	13.7	3531	4	US-09-169-768-7	Sequence 7, Appl1
41	263.5	13.7	3541	4	US-09-169-768-9	Sequence 9, Appl1
42	263.5	13.7	4409	4	US-09-331-347C-22	Sequence 22, Appl
43	263	13.7	6200	4	US-09-795-061-3	Sequence 3, Appl1
44	262.5	13.7	22218	4	US-09-949-016-14240	Sequence 14240, A
45	261.5	13.6	2070	4	US-09-949-016-5405	Sequence 5405, Ap

ALIGNMENTS

RESULT 1
US-08-343-443B-1
; Sequence 1, Application US/08343443B
; Patent No. 5968734
; GENERAL INFORMATION:
; APPLICANT: Aurias, Alain
; APPLICANT: Delattre, Olivier
; APPLICANT: Desmazes, Chantal
; APPLICANT: Melot, Thomas
; APPLICANT: Peter, Martine
; APPLICANT: Ploogastel, Beatrice
; APPLICANT: Thomas, Gilles
; APPLICANT: Zuchman, Jessica
; TITLE OF INVENTION: NUCLEIC ACID CORRESPONDING TO A GENE OF
; TITLE OF INVENTION: CHROMOSOME 22 INVOLVED IN RECURRENT CHROMOSOMAL
; TITLE OF INVENTION: TRANSLOCATIONS ASSOCIATED WITH THE DEVELOPMENT OF CANCEROUS
; TITLE OF INVENTION: TUMORS AND NUCLEIC ACIDS OF FUSION RESULTING FROM SAID
; TITLE OF INVENTION: TRANSLOCATIONS
; NUMBER OF SEQUENCES: 129
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Weiser & Associates
; STREET: 230 South Fifteenth Street
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: AEDIT 1.0 DOS text editor
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/343,443B
; FILING DATE: 18-NOV-1994
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/FR93/00494
 FILING DATE: 19-MAY-1993
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: FR 92/06123
 FILING DATE: 20-MAY-1992
 ATTORNEY/AGENT INFORMATION:
 NAME: Weisner, Gerard J.
 REGISTRATION NUMBER: 19,763
 REFERENCE/DOCKET NUMBER: 989,6121P
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 215-875-8383
 TELEFAX: 215-875-8394
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2371 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 25..1992
 US-08-343-443B-1

Alignment Scores:
 Pred. No.: 3,06e-97 Length: 2371
 Score: 1922.00 Matches: 338
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 Gaps: 0

US-10-791-017a-2_COPY_319_656 (1-338) x US-08-343-443B-1 (1-2371)

QY 1 G1YGLVARG1YGLYMETG1YSER1AG1YGLUARG1YGLYPHEANLYSPROGLYGLY 20
 DB 979 GGAGAGCGCGGTGAAATGGGCGCGTGGAGCGAGCGAGTGGCTTCAATAAGCCGTGGGA 1038
 QY 21 PROMETAPG1UG1YPROAPLEUAPLEUG1YPROPROVALAPPROAPG1UAPSER 40
 DB 1039 CCCATGATGAGGACGACGAGCTTGAATCGGCTCTCTAGATCCAGATGAAGACTCT 1098
 QY 41 ASPANSEALAIETYRVALG1NG1YLEUASAPSERVALTHREUASAPLEU1A 60
 DB 1099 GACAAAGCTGCAATTATGTCACAGATTAATGACAGTGTGACTTACATGATCTGCA 1158
 QY 61 ASPHEPHELYSG1NCYSG1YVALVALYMECANLYSARQTHRG1YGLINPROMET1LE 80
 DB 1159 GACTTCTTTAAGCAGTGTGGGTGTGTTAAGATGAACAAGAGACTGGGCAACCATGATC 1218
 QY 81 H1A1ETYLEUAPLEUAPLEUG1YMETG1YLEUAPROLYSG1YASAP1ATHRVALSER1YRGLU 100
 DB 1219 CACATCTACCTGGACAGGAAAGGAAAGCCCAAGGCGATGACACAGTGTCTTATGAA 1278
 QY 101 ASPPROBOTHRALALYVALA1A1VAL1G1UTRPHASPG1YLYASAPHEGL1YSER 120
 DB 1279 GACCAACCCACTGCGCAAGCGCTCGGTGAATGTTTATGGAAGAAATTTTCAAGGAGC 1338
 QY 121 LYSLEULYVALSERLEUALATRYLYLYBPROBOMETCANSEMERCHRG1YGLYLEU 140
 DB 1339 AAACCTTAAAGTCTCCCTGCTCGGAGAGACCTCAATGAAACATATGCGGGGTGCTCG 1398
 QY 141 PROPROARG1UG1YARG1YMETPROBROBOMETARG1YGLYPRORGLYGLY 160
 DB 1399 CCACCCCGTGAAGGAGAGGATGCGACACACCTCGGTGAAGTCCAGGAGGCCCAAGGA 1458
 QY 161 G1YPROGLYGLYPROMETG1YARGMETG1YGLYARG1YGLYASPARG1YGLYPHAPRO 180
 DB 1459 GGTCTCTGGGAGCCCATGGGTGCGATGGAGGCCCTGTGAGAGATTAAGAGGCTTCCCT 1518
 QY 181 PROARG1YPROARG1YSEARGLYASANDPROSEARG1YGLYLAENVALG1NH1SARG 200
 DB 1519 CCAAGAGGACCCCGGGGTTCCTCGAGAGGAAACCTCTGTGAGAGGAAACGTCTCAACACGA 1578

QY 201 A1AG1YASPTIRG1NCYSPROBANPROGLYCYSG1YASNG1NAPSEALATPRARG1THR 220
 DB 1579 GCTGAGACTGSCATGTCTCCCAATCCGGTTCTGGAACCAACTCCGCTGGAGACA 1638
 QY 221 GIUCYASNG1NCYLYVALAPROLYSPROGLYGLYPHELEUPROBROBOPRO 240
 DB 1639 GAGTCAACCACTGTAAGGCCCAAGCCTGAAGGCTTCTCCGCCACCTTCCGCCCC 1698
 QY 241 PROGLYGLYASPARG1YARG1YGLYPROGLYGLYMETARG1YGLYARG1YGLYLEU 260
 DB 1699 CCGGGGTGATGTCGCGAGAGTGGCCCTGTGTCATGCGGAGAGAGAGTGGCTTC 1758
 QY 261 METASPARG1YGLYPROGLYGLYMETPHEARG1YGLYARG1YGLYASPARG1YGLY 280
 DB 1759 ATGATCGTGTGTCGTCTCCGTGGAATGTTCAAGGTGGCGCGTGTGAGACAGAGTGGC 1818
 QY 281 PHEARG1YGLYARG1YMETASPARG1YGLYPHEGLYGLYARG1YGLYARG1YGLYPRO 300
 DB 1819 TTCGTGTGTGGCGGGGCATGACCGAGCTGTGCTTGTGTGTGAGAAAGAGTGGCCCT 1878
 QY 301 G1YGLYPROPROGLYPROLEUMETG1UG1METG1YGLYARG1YGLYARG1YGLY 320
 DB 1879 GGGGGGCCCCCTGGACCTTGTGATGAAACATGGGAGAGAAAGAGAGACGTGAGCA 1938
 QY 321 PROGLYLYMETASPLYSGLYGLUHSARG1NG1UARARG1YGLYASPARPROTYR 338
 DB 1939 CCTGGAATAATGATTAAGCCAGACCGCTCAGAGCGCAGATGGGCCCTAC 1992

RESULT 2
 US-09-949-016-5043
 Sequence 5043, Application US/09949016
 Patent No. 6812339
 GENERAL INFORMATION:
 APPLICANT: VENTER, J. Craig et al.
 TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
 TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
 FILE REFERENCE: CL001307
 CURRENT APPLICATION NUMBER: US/09/949,016
 PRIOR FILING DATE: 2000-04-14
 PRIOR APPLICATION NUMBER: 60/241,755
 PRIOR FILING DATE: 2000-10-20
 PRIOR APPLICATION NUMBER: 60/237,768
 PRIOR FILING DATE: 2000-10-03
 PRIOR APPLICATION NUMBER: 60/231,498
 PRIOR FILING DATE: 2000-09-08
 NUMBER OF SEQ ID NOS: 207012
 SOFTWARE: FASTSEQ for Windows Version 4.0
 SEQ ID NO 5043
 LENGTH: 1785
 TYPE: DNA
 ORGANISM: Human
 US-09-949-016-5043

Alignment Scores:
 Pred. No.: 3.25e-91 Length: 1785
 Score: 1810.00 Matches: 318
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 94.17% Indels: 0
 Gaps: 0

US-10-791-017a-2_COPY_319_656 (1-338) x US-09-949-016-5043 (1-1785)

QY 21 PROMETASPG1UG1YPROAPLEUAPLEUG1YPROPROVALAPPROAPG1UAPSER 40
 DB 829 CCCATGATGAGAGACCAAGATCTTGTATCTAGGCCCACTGTGATCCAGATGAAGCTCT 888
 QY 41 ASPANSEALAIETYRVALG1NG1YLEUASAPSERVALTHREUASAPLEU1A 60
 DB 889 GACAAAGCTGCAATTATGTCACAGATTAATGACAGTGTGACTTACATGATCTGCA 948
 QY 61 ASPHEPHELYSG1NCYSG1YVALVALYMECANLYSARQTHRG1YGLINPROMET1LE 80

```
Db      949 GACTCTTTAAGCAGTGGGGTTGTTAAGTGAACAAGAACTGGGCAACCATGATC 1008
Qy      81 Hs11eTyleuAplysGIuThrGIyLysProlyseGIyAspa1aThrValSerTyrgIu 100
Db      1009 CACATCTACCTGACAAAGAAACAGAAAGCCCAAGGCGATGCCACAGTCTCTATGAA 1068
Qy      101 AspProThralAlaYsa1a1aValaGIuThrPheaspGIyLysAspPheGInglySer 120
Db      1069 GACCCACCACTGCAAGGCTCCGTAAGATGGTTGATGGAAAGATTTTCAAGGGAGC 1128
Qy      121 LysLeuLysValSerleu1aAargIyLysProPheAasnerMetAargIyGIyLeu 140
Db      1129 AAACCTTAAGTCTCCCTGCTCGGAAGAAGCTCCATGAACGTAAGCGGGGTGGCTG 1188
Qy      141 ProPheAargIyGIyLysMetProPheProleuAargIyGIyProGIyGIyProGIy 160
Db      1189 CCACCCCGTGAAGGCGAGAGCATGCCACCACTCGTGAAGTCCAGAGAGCCCGCAGGA 1248
Qy      161 GIyProGIyGIyProMetGIyAargMetGIyGIyAargIyGIyLysAspAargGIyLysPhePro 180
Db      1249 GGTCTCGGGGAGCCCAATGGATGGATGGAGCCGTGAGAGATAGAGAGAGCTTCCCT 1308
Qy      181 ProAargIyProAargIySerAargIyAasProSerGIyGIyGIyAasValaGIyH1aarg 200
Db      1309 CCAAGAGAGACCCCGGGGTTCCCAAGGGAACCTCTTGAGAGAGAAACCTCCAGCAGCA 1368
Qy      201 AlaGIyAspTrpGIyGIyLysProAasProGIyGIyGIyAasGInaAspPhe1a1aTrpAargThr 220
Db      1369 GCTGGAGACTGCGAGTGTCCCAATCCGGGTGTAAGAAACAGAACTTCGCTGAGAGACA 1428
Qy      221 GIuCyAasGInaGIyLysA1aProlYsPProGIyGIyPheLeuProPheProPro 240
Db      1429 GAATGCAACCAAGTAAAGGCCCAAGGCTGAAGGCTTCTCCGCAACCTTTCGCGCC 1488
Qy      241 ProGIyGIyAAspAargIyAargIyGIyProGIyGIyLysMetAargIyGIyAargIyGIyLeu 260
Db      1489 CCGGGTGTGATCTGTGCAAGAGTGGCCCTGTGTGCAATGGGGAGAGAGAGTGGCTC 1548
Qy      261 MetAspAargIyGIyLysProGIyGIyLysMetPheAargIyGIyAargIyGIyAAspAargIyGIy 280
Db      1549 ATGATCGTGTGTGTCCCGGTGAATGTTCAGAGGTGGCCGTGTGTGAACAAGAGTGGC 1608
Qy      281 PheAargIyGIyAargIyLysMetAAspAargIyGIyLysPheGIyGIyLysAargIyGIyPro 300
Db      1609 TTCCGTGTGTGCGCGGAGCATGAGCCAGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1668
Qy      301 GIyGIyProProGIyProleuMetGIuGInMetGIyGIyAargAargGIyGIyAargGIyGIy 320
Db      1669 GGGGGGCCCCCTGACCTTTGATGGAACAGATGGAGAGAAAGAGAGAGAGAGAGAGAGAGAGAG 1728
Qy      321 ProGIyLysMetAAspIyGIyGIyLysAargGInaAargAargAargAargAargAargAargAarg 338
Db      1729 CCGGAAAAAATGATTAAGCGAGACCGTCAAGAGCGAGAGATGGCGCTTAC 1782

RESULT 3
US-09-949-016-5044
; Sequence 5044, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: C1001307
; CURRENT APPLICATION NUMBER: US/09/949, 016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
```

```
; SEQ ID NO 5044
; LENGTH: 1785
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-5044

Alignment Scores:
Pred. No.: 3,25e-91 Length: 1785
Score: 1810.00 Matches: 318
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 94.17% Indels: 0
DB: 4 Gaps: 0

US-10-791-017a-2_COPY_319_656 (1-338) x US-09-949-016-5044 (1-1785)

Qy      21 ProMetAAspGIuGIyProAAspLeuAAspLeuGIyProProValAAspProAAspSer 40
Db      829 CCCATGGATGAAGACCAAGATCTTGATCTTAGGCCCACTGTATGATCCAGATGAAGACTCT 888
Qy      41 AspaSner1a1eTyrValaGIyLysAasPserValThreAAspAAspLeu1a 60
Db      889 GACAAAGTGCATTTATGTAACAAGATTAAATGACAGTGTGACTTACATGATCTTGCA 948
Qy      61 AspPhePheLysGInaGIyLysValaLysMetAAspLysAargThrGIyGInProMetIle 80
Db      949 GACTCTTTAAGCAGTGGGGTTGTTAAGATGAACAAGAACTGGGCAACCATGATC 1008
Qy      81 Hs11eTyleuAplysGIuThrGIyLysProlyseGIyAspa1aThrValSerTyrgIu 100
Db      1009 CACATCTACCTGACAAAGAAACAGAAAGCCCAAGGCGATGCCACAGTCTCTATGAA 1068
Qy      101 AspProThralAlaYsa1a1aValaGIuThrPheaspGIyLysAspPheGInglySer 120
Db      1069 GACCCACCACTGCAAGGCTCCGTAAGATGGTTGATGGAAAGATTTTCAAGGGAGC 1128
Qy      121 LysLeuLysValSerleu1aAargIyLysProPheAasnerMetAargIyGIyLeu 140
Db      1129 AAACCTTAAGTCTCCCTGCTCGGAAGAAGCTCCATGAACGTAAGCGGGGTGGCTG 1188
Qy      141 ProPheAargIyGIyLysMetProPheProleuAargIyGIyProGIyGIyProGIy 160
Db      1189 CCACCCCGTGAAGGCGAGAGCATGCCACCACTCGTGAAGTCCAGAGAGCCCGCAGGA 1248
Qy      161 GIyProGIyGIyProMetGIyAargMetGIyGIyAargIyGIyLysAspAargIyGIyPro 180
Db      1249 GGTCTCGGGGAGCCCAATGGATGGATGGAGCCGTGAGAGATAGAGAGAGCTTCCCT 1308
Qy      181 ProAargIyProAargIySerAargIyAasProSerGIyGIyGIyAasValaGIyH1aarg 200
Db      1309 CCAAGAGAGACCCCGGGGTTCCCAAGGGAACCTCTTGAGAGAGAAACGTCAGCAGCA 1368
Qy      201 AlaGIyAspTrpGIyGIyLysProAasProGIyGIyGIyAasGInaAspPhe1a1aTrpAargThr 220
Db      1369 GCTGGAGACTGCGAGTGTCCCAATCCGGGTGTAAGAAACAGAACTTCGCTGAGAGACA 1428
Qy      241 ProGIyGIyAAspAargIyAargIyGIyProGIyGIyLysMetAargIyGIyAargIyGIyLeu 260
Db      1489 CCGGGTGTGATCTGTGCAAGAGTGGCCCTGTGTGCAATGGGGAGAGAGAGTGGCTC 1548
Qy      261 MetAspAargIyGIyLysProGIyGIyLysMetPheAargIyGIyAargIyGIyAAspAargIyGIy 280
Db      1549 ATGATCGTGTGTGTCCCGGTGAATGTTCAGAGGTGGCCGTGTGTGAACAAGAGTGGC 1608
Qy      281 PheAargIyGIyAargIyLysMetAAspAargIyGIyLysPheGIyGIyLysAargIyGIyPro 300
Db      1609 TTCCGTGTGTGCGCGGAGCATGAGCCAGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1668
Qy      301 GIyGIyProProGIyProleuMetGIuGInMetGIyGIyAargAargGIyGIyAargGIyGIy 320
```



```

Db      1669 GGGGGGGCCCCCTGACCTTTGATGCAACAGATGGAGGAAAGAGAGACCTGAGCA 1728
Qy      321 ProGlyLysMetAspLysGlyLysIleAspGlnGluArgArgAspArgProTy 338
Db      1729 CTTGGAAAAATGATTAAGGCGAGCACCTGACAGAGCGAGAGATCGCCCTAC 1782

RESULT 4
US-09-949-016-13696
; Sequence 13696, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTNER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ. ID NOS: 207012
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 13696
; LENGTH: 6002
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-13696

Alignment Scores:
Pred. No.: 6,14e-81 Length: 6002
Score: 1632.00 Matches: 299
Percent Similarity: 91.72% Conservative: 11
Best Local Similarity: 88.46% Mismatches: 28
Query Match: 84.91% Indels: 2
DB: 4 Gaps: 0

US-10-791-017a-2_COPY_319_656 (1-338) x US-09-949-016-13696 (1-6002)
Qy      1 G1G1YARGLYRG1YMETGLYSERIALG1GLUARG1YGLYPHEANLYSPROGLYGLY 20
Db      2988 GAGAGAGCGCGTGGATAGGCGAGCGCTGGAGAGCAAGTGGCTTCAATTAAGCTCGTGA 3047
Qy      21 ProMetAspGluGlyProAspLeuAspLeuGlyProProValAspProAspGluAspSer 40
Db      3048 CCCATGGATGAGAGCCAGATCTTGATCTAGGCCCATCTGTAGATCCAGATGAAGACTCT 3107
Qy      41 AspAsnSerAlaIleTyValGlnGlyLeuAsnAspSerValThrLeuAspAspLeuAla 60
Db      3108 GACAAACGTCGAATTTATGTACAAAGATTAATACATGACTCTACATGATGTGTA 3167
Qy      61 AspPhePheLysGlnGlyValValIleMetAsnLysAspThrGlyGlnProMetIle 80
Db      3168 GACTCTTTAAGCAGTGTGGGCTTTTAAGTAAAGTAAAGACCAAGCACTGAGCAACCATATC 3227
Qy      81 HisIleTyLeuAspLysGlnGlyValIlePheProLysGlyAspAlaThrValSerTyGlu 100
Db      3228 CACACCTAAGCTGACCAAGAAACAAGAACCCAAAGGTGATGCCACAGTCTCTGTGAA 3287
Qy      101 AspProProThrAlaValAlaValAlaValGluThrPheAspGlyLysAspPheGlnGlySer 120
Db      3288 GACTCACCTAAGCTGCAAGAGCTCGGTGAAATGTTTGAATGGAAAGATTTTCAAGGAGC 3347
Qy      121 LysLeuLysValSerLeuAlaArgLysLysProPheMetAsnSerMetArgGlyGlyLeu 140
Db      3348 AAACTTAAAGTCTCTTCTCTCGAAGAGGCTCTCAAGTAAAGTAAAGTAAAGTAAAGTAA 3407
Qy      141 ProProArgGluGlyArgGlyMetProProProLeuArgGlyGlyProGlyGlyProGly 160
Db      3408 CCACCCGTAAGGCGAGAGGATGTCACCACTCTGCGAGAGGTCCAGAGAGGCCCAAGA 3467

```

```

Qy      161 G1YProGlyLysProMetGlyArgMetGlyLysArgGlyLysAspArgGlyLysPhePro 180
Db      3468 AGTCTCGAGGAGCCCATGGTCAATGGAGGCGCTGGAGAGATGAGAGAGCCCTCCCT 3527
Qy      181 ProArgGlyProArgGlySerArgLysAsnProSerGlyGlyLysValGlnIleAspArg 200
Db      3528 CCAAGAGAGCCCAAGGTTCCCGAGAGAACCTCTGAGAGAGAAACCTCCAGCACCA 3587
Qy      201 AlaGlyAspTrpGlnCysProAsnProGlyCysGlyValAsnGlnAsnPheAlaTrpAspThr 220
Db      3588 GCTGGAGACAGCAGTGTCCCAATCGGCTTGTGAACCAAGACTTCCCTGGAGAAC 3647
Qy      221 G1uCyAsnGlnCysLeuAlaProLysProGlnGlyPheLeuProProPheProPro 240
Db      3648 GAGAGACAGAGTAAAGCTTCAAGCTTAAAGCTTCTCCGACACCTTCCACCC 3707
Qy      241 ProGlyLysAspArgGlyArgGlyLysProGlyLysMetArgGlyLysArgGlyLysLeu 260
Db      3708 CCGGATGATCATGATGAGAGAGTGGCTCTGTGGCATGTGGGAGAGAGAGTGGCTTC 3767
Qy      261 MetAspArgGlyLysProGlyLysMetPheArgGlyLysArgGlyLysAspArgGlyLys 280
Db      3768 ATGATCATGATGATGCTCCGCTGAAATGTTCAAGAGTGTGTGTGAGACAGAACTGCTC 3827
Qy      281 PheArgGlyLysArgLysMetAspArgGlyLysPheGlyGlyLysArgArgGlyLysPro 300
Db      3828 TTCCTGTGTGCTGCGCATGACCGAGTGGCTTTGTGGAGAGAGACAGAGTGGCTCT 3886
Qy      301 G1GlyLysProGlyLysProLeuMetGlnGlnMetGlyLysArgArgGlyLysArgGlyLys 320
Db      3887 GGGGGGGCCCCCGGA-CCTTTGATGTAAACAATGGAGAGAAAGAGAGAGAGAGAGAG 3945
Qy      321 ProGlyLysMetAspLysGlyLysIleAspGlnGluArgArgAspArgProTy 338
Db      3946 CTTGGAAAAACGATTAAGCGAGCAGCTGTCAAGAGCGAGAGATCAGCCCTAC 3999

RESULT 5
US-09-949-016-1954
; Sequence 1954, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTNER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ. ID NOS: 207012
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 1954
; LENGTH: 1783
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-1954

Alignment Scores:
Pred. No.: 1.87e-76 Length: 1783
Score: 1541.00 Matches: 282
Percent Similarity: 91.82% Conservative: 10
Best Local Similarity: 88.68% Mismatches: 26
Query Match: 80.18% Indels: 2
DB: 4 Gaps: 0

US-10-791-017a-2_COPY_319_656 (1-338) x US-09-949-016-1954 (1-1783)
Qy      21 ProMetAspGluGlyProAspLeuAspLeuGlyProProValAspProAspGluAspSer 40

```


QY	154	-----	154
Db	32106	GGAGAAATTGTTGAATCTGGGGGGGTGGAGTTGCAGTGAACAAAGTGTGCACCTGCA	321655
QY	154	-----	154
Db	32166	CTCCAGCCTGGCAACAGTGTGAGACTTCGCTCTCAAAAAAAAAAAAAAAAAATTGTGGG	322255
QY	154	-----	154
Db	32226	AGCTCTGTTCTGTAGAGACGTGGAAACGCTCTCCACAGGAGAGGGGGCTGATGCTCT	322855
QY	154	-----	154
Db	32286	GAGCCACACGGAACACGGGACAGGTGTGGGAAAAATGACAGAGTATCTGTGGCTT	323455
QY	154	-----	154
Db	32346	TACTTAGTATTTTATTTCTCTATAGCAAAATTGGTGTCTACAGAGAAATGATTTGCTGT	324055
QY	155	-----	155
Db	32406	TCCTGTGTTCTGTGTAGGTCCAGAGAGGCCAGAGGTCCTGGGGGACCCATGGGGTGG	324655
QY	168	gMecllygllyAarglygllyAaParglygllyPheProProArgglyProArglyPheArg	168
Db	32466	CATGGAGAGCCGTGGAGAGATGACAGAGCTTCCCTCCAGAGAGACCCGGGGTTCCGG	325255
QY	188	gGlyAaPProSerglygllyAaenValGlnHisArgAlaGlyAaPTrpGlnCyAProAs	208
Db	32526	AGGGAACCCCTCTGGAGGAGAAACGTCCAGACACCGAGCTGGAGACTGTGCCAA	325855
QY	208	nProGly-----	210
Db	32586	TCCGTA-TGTACTGTCTTGGCAAAATTGATACCTTAGAGTGAAGCCACCTTCCCTCAC	326445
QY	210	-----	210
Db	32645	CCCATCCCACTTAGAGTGAATGCTGTCTGTCTAGAGAAACAGATGATGACCTGATGG	327045
QY	210	-----	210
Db	32705	CTGTTAGGAGACACTAGTCAAGCCATTGACGTGACGTTCAAGGCTTCTGAAGATTGATT	327645
QY	210	-----	210
Db	32765	TGACCTGTCTGTGGTGCAATGCTGCTGAGCTGTGCTTAAGCANTGGTGTACATA	328245
QY	210	-----	210
Db	32825	GATCCTCTTGATAGTGAAGTGTGATCCTGTTCACACACCACTTCTCTGTTATCTTCTCT	328845
QY	211	-----	215
Db	32885	TAGTTGAAATGGGTATTTCTGTCTGTGATGATGATATGATGACAGGGGTGTGGAAACAGAA	329445
QY	215	nPheAlaTrpArgThrGluCyAaenGlnCyAaenAlaProlyAaProlyGluGlyPheLeuP	235
Db	32945	CTTCCGCTCGAGAAACAGAGTGCACACAGTGTAAAGCCCAAGCCTGAAAGCTTCCCTCC	330045
QY	235	oProProPheProProPro-----	241
Db	33005	GCCACCCCTTTCGCCCCCGGGGTAGGTGCAAGTTTCATGATGTCCCTCAGCTTCCCTGGT	330645
QY	241	-----	241
Db	33065	GCTAAACCTCTTTTCTTATTTGTGGGCTTGGTAAACGTCACTGCCTCTGCTTAACAC	331245
QY	241	-----	241
Db	33125	TTTGAGTTGTGTGTCTCATTTCTTAAATTGTGACGCCGAGTCCGAGATTGAGTGAAGTG	331845
QY	241	-----	241

```

Db      33185 TCTGTGTTGTTCTGTGTGAGAGAAAGAAAGCAGACAGTTCACAGTGTCCACAGGGCC 33244
QY      242 -----GlyGlyAspArgI 246
Db      33245 TCTGCAGCCACCACTGACTGCTTGCCTGTAATTCACCTTAGTGATGTGTGG 33304
QY      246 YArgGlyGlyProGlyGlyMetArgGlyGlyYArgGlyGlyLeuMetAspArgGlyGlyPr 266
Db      33305 CAGAGGTGCGCCCTGTGTGCAATGCGGGAGAAAGAGTGGCTCATGAGATCGATGTGTCC 33364
QY      266 OGlyGlyMetPheArgGlyGlyYArgGlyGlyYAspArgGlyGlyPheArgGlyGlyYArgG 286
Db      33365 CGGTGATGATGTTCAAGAGGTGGCCGGGTGGAGACAGAGGTGGCTTCCGTGTGTGGCCGGGG 33424
QY      286 YMetAspArgGlyGlyYPheGlyGlyGlyYArgArgGlyGlyGlyProGlyGlyProGlyPr 306
Db      33425 CATGACCCGAGAGTGGCTTTGTGTGGAGGAAGACAGAGTGGCCCTTGGGGGGCCCCCTGGACC 33484
QY      306 OLeuMetGlyGlyInMetGlyGlyYArgArgGlyGlyYArgGlyGlyProGlyYLeuMetAsp 326
Db      33485 TTGATGTAACAGATGGAGAGAAAGAAAGAGAGACTGTGAGGACCTCGAAAATATGATTA 33544
QY      326 B----- 326
Db      33545 -GTAAAGTCGTGTGTAAGAAAGCAGCTGTGGAGCGCCAGGCAAGTAAGAGACAGCCCTTCC 33603
QY      326 ----- 326
Db      33604 CAGCTGTGTTGGCCGCAAGTCTCATGTCTTAGAAGACTTGTGATAGTGTGGAGAGAG 33663
QY      326 ----- 326
Db      33664 CCAGAGAGGGGACACTGGGGGCTTGAAAGGGCTTCTCAACCCCTTCCATTAACGA 33723
QY      327 -----GlyGlyHisArgGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 338
Db      33724 AGGGCCCTTTTACTTTCAGCTTGCAGAGGGAGACACCGTCAGAGCGCAGAGATGGCCCTAC 33781

RESULT 7
US-09-949-016-16786
Sequence 16786, Application US/09949016
Patent No. 6812339
GENERAL INFORMATION:
APPLICANT: VENTER, J. Craig et al.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
FILE REFERENCE: C1001307
CURRENT APPLICATION NUMBER: US/09/949,016
CURRENT FILING DATE: 2000-04-14
PRIOR APPLICATION NUMBER: 60/241,755
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/237,768
PRIOR FILING DATE: 2000-10-03
PRIOR APPLICATION NUMBER: 60/231,498
PRIOR FILING DATE: 2000-09-08
NUMBER OF SEQ ID NOS: 207012
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 16786
LENGTH: 35784
TYPE: DNA
ORGANISM: Human
US-09-949-016-16786

Alignment Scores:
Pred. No.: 3,336-43 Length: 35784
Score: 958.00 Matches: 233
Percent Similarity: 29.75% Conservative: 5
Best Local Similarity: 29.12% Mismatches: 11
Query Match: 49.84% Indels: 554
DB: 4 Gaps: 6

US-10-791-017A-2_COPY_319_656 (1-338) x US-09-949-016-16786 (1-35784)

```



```

Db      33545  GTAAAGCTGCTGTGTAATAAGCAGCTGTGGCCGCGCAGGCA CAGTAAAGGACACGCTTCC 33603
Oy      326  -----326
Db      33604  CAGCTGTGTGGCGCAAGTCTCATGTCCTAGGAACCTTGATGATGTTGGAGAG 33663
Oy      326  -----326
Db      33664  CCAGAAAGGGACACTGTGGGGCTCTGGAAGAGGCTTCTCAACCCCTTCCATTCGAACGA 33723
Oy      327  -----GlyIuH!sAtgGIngluAArgAAspArgProTyr 338
Db      33724  AGGGCCCTCTTAACCTTGACAGAGCGAGCACCGCTCAGAGACGAGAGATGCGCCCTAC 33781
RESULT 8
US-09-919-039-322
/ Sequence 322, Application US/09919039
/ Patent No. 6727066
/ GENERAL INFORMATION:
/ APPLICANT: Kasert, Matthew R.
/ TITLE OF INVENTION: GENES EXPRESSED IN TREATED HUMAN C3A LIVER CELL CULTURES
/ FILE REFERENCE: PA-0035 US
/ CURRENT APPLICATION NUMBER: US/09/919,039
/ CURRENT FILING DATE: 2002-09-09
/ PRIOR APPLICATION NUMBER: 60/222,113
/ PRIOR FILING DATE: 2000-07-28
/ NUMBER OF SEQ ID NOS: 401
/ SOFTWARE: PERL Program
/ SEQ ID NO 322
/ LENGTH: 1939
/ TYPE: DNA
/ ORGANISM: Homo sapiens
/ FEATURE:
/ NAME/KEY: misc feature
/ OTHER INFORMATION: Incyte ID No. 6727066 478620.53
US-09-919-039-322

```

Alignment Scores:		4.9e-40	Length:	1939
Pred. No.:		678.00	Matches:	182
Score:		63.37%	Conservative:	36
Percent Similarity:		52.91%	Mismatch:	54
Best Local Similarity:		45.68%	Indels:	72
Query Match:		4	Gaps:	13
DB:				

QY	US-10-791-017A-2_COPY_319_656 (1-338) x US-09-919-039-322 (1-1939)	
QY	1 G1YGYAAGG1YGLMwG1YSer1aG1YwAGG1YGLYpHAsu1YpR0G1YGL	20
DB	825 GGAGGACAGGGGCGCATGGCCGAAGT--GACCGTGGGCTTAATTAATTTGTGGC	881
QY	21 PwMeAspG1YGLYpRAspLeuAspLeuG1YpR0pR0ValAspP0AspG1YwAspSer	40
DB	882 CCTCGGGACCAAGAGATCAAGTCATGAC-----TCCGAACAGGATATATCA	926
QY	41 AspaAsSer1a1e1YtYValG1nG1YLeuAsnAspSerVal1Th1LeuAspAspLeu1a	60
DB	927 GACCAACAAACCATTTGTGGCAAGCCCTGGGTGAAGATGTACATTAAGTCGTGGCT	986
QY	61 AspPheLeuYsG1nCyseG1YValVal1YsMe1aen1YsArgThrG1YGLnPrMe11e	80
DB	987 GATTACTTCAAGCAATAGTGATTTATTTTAAGACAAACAAAGAAACGGACAGCCATGATT	1046
QY	81 H1s1e1eYTLLeuAsp1YsG1nThrG1YLYsP0d1YsAsp1a1Th1ValSer1Yc1u	100
DB	1047 AATTGTACACAGACAGGAAACTGGCAAGCTGAAGGAGAGACCAAGCTCTCTTTGAT	1106
QY	101 AspP0pR0Th1a1e1Ys1a1a1a1aValG1n1uTPR1PheAspG1YLYAspPheG1nG1YsEr	120
DB	1107 GACCAACCTTCAGCTTAAGACAGCTATTGACTGTGTTTGATGTGTAAGTAATTCGCGGAAT	1166
QY	121 Lyb1e1uYsValSer1e1u1aArg1Ys1YpP0PwMe1aenSerMe1Arg1YGLYLeu	140

Db	1167	CCATTCAAGGCTTCATTGGTACTCCGCGGAGACTTTAT-----CGGAGGTGGTGC	1220
QY	141	ProProArgGluGlyArgGlyMetProProProLeuAlaArgGlyGlyProGlyGlyProGly	160
Db	1221	AATGGTGTGGAGGCGCAGGG-----CGAAGAGACCAAGGCGCGTGGGA	1265
QY	161	GlyProGlyGlyProMetGlyArgMetGlyGlyArgGlyGlyAspArgGlyGlyPhePro	180
Db	1266	GGCTATGAGGTGGT-----GGCAGTGTGTGTGTGGCCGAGGAGATTCCC	1313
QY	181	ProArgGlyProAlaArgGlySerArgGlyAsnProSerGlyGlyGlyAsnValGlnHisArg	200
Db	1314	AGTGAAGT-----GTGGCGGTGGAGACAGCACGGA	1346
QY	201	AlaGlyAspTrpGlnCysProAsnProGlyCysGlyAsnGlnAsnPheAlaTrpArgThr	220
Db	1347	GCTGTGATCTGGAAGTCTCTAATCCACCTGTGAGAAATGAACCTCTTGGAGGAAT	1406
QY	221	GluCysAsnGlnCysValAlaProLysProGlnGlyPheLeuProProPheProPro	240
Db	1407	GAATGCAACCAAGTGTAAAGGCCCTTAAACGAGATGC-----	1442
QY	241	ProGlyGlyAspArgGlyArgGlyGlyProGlyGlyMetArg---GlyGlyArgGlyGly	259
Db	1443	CCAGAA-----GGGGGACAGGTGGCTCTCACTGGGGGTAACTACGGG	1487
QY	260	LeuMetAspArgGlyGlyProGlyGlyMetPheArgGly-----GlyArgGlyGly	276
Db	1488	GATGATCGTGTGTGGCAGAGGAGGCTATGATGAGAGCGGCTACCGGGGCGGGCGGG	1547
QY	277	AspArgGlyGlyPheAspArgGlyGlyArg---GlyMetAspArgGlyGlyPheGlyGlyGly	295
Db	1548	GACCTGTAGAGCTTCCAGAGGGGCGGGGGTGGGGACAAAGGTGGCTTT-----	1598
QY	296	ArgArgGlyGlyProGlyGlyProProGlyProLeuMetGlnGlnMetGlyGlyArgArg	315
Db	1598	-----	1598
QY	316	GlyGlyArgGlyGlyProGlyGlyMetAsp---LysGlyGlyHisArgGlnGluArgArg	334
Db	1599	-----GGCCCTGGCAATGATGATTCCAGGGGTGAGCACAAGAGATCGCAGG	1646
QY	335	AspArgProTyr	338
Db	1647	GAGAGGCCGTTT	1658

```

RESULT 9
US-09-621-976-13361
; Sequence 13361, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 13361
; LENGTH: 411
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-621-976-13361

Alignment Scores:
Pred. No.:      6.33e-26
Score:          609.00
Percent Similarity: 99.15%
Best Local Similarity: 99.15%
Query Match:    31.69%

Length: 411
Matches: 116
Conservative: 0
Mismatch: 0
Indels: 1

```



```
QY 326 ----- 326
Db 241 GCCGCCAGCAGACAGTAGAGACAGCCCTTCCAGCTTGTTGGCGAAGTCTCATGTC 300
QY 326 ----- 326
Db 301 KCTAGAGACTTGTGATAGTGGTTGGAGAGAGCCAGAGAGGACCTGGGGGCTCTGGA 360
QY 327 ----- 327
Db 361 AGGGCTCTCTCAACCCCTTCCATTTCAACGAGAGGCGCTTTTACCTTGACAGAGCGAG 420
QY 329 H1sargIngluAargAargAparProtyr 338
Db 421 CACCGTCAGAGAGCGCAGAGTCCGCCCTTAC 450

RESULT 12
US-09-370-838-145
/ Sequence 145, Application US/09370838
/ Patent No. 6444425
/ GENERAL INFORMATION:
/ APPLICANT: Reed, Steven G.
/ APPLICANT: Lodes, Michael J.
/ APPLICANT: Mohamath, Raodoh
/ APPLICANT: Secrist, Heather
/ TITLE OF INVENTION: COMPOUNDS FOR THERAPY AND DIAGNOSIS OF
/ TITLE OF INVENTION: LUNG CANCER AND METHODS FOR THEIR USE
/ FILE REFERENCE: 210121.475C1
/ CURRENT APPLICATION NUMBER: US/09/370,838
/ EARLIER FILING DATE: 1999-08-09
/ EARLIER APPLICATION NUMBER: US 09/285,323
/ NUMBER OF SEQ ID NOS: 289
/ SOFTWARE: FaastSeq for Windows Version 3.0
/ SEQ ID NO 145
/ LENGTH: 450
/ TYPE: DNA
/ ORGANISM: Homo sapien
US-09-370-838-145

Alignment Scores:
Pred. No.: 2,4e-11 Length: 450
Score: 344.00 Matches: 64
Percent Similarity: 79.468 Conservative: 25
Best Local Similarity: 57.144 Mismatches: 16
Query Match: 17.908 Indels: 7
Gaps: 2

US-10-791-017a-2_COPY_319_656 (1-338) x US-09-370-838-145 (1-450)
QY 1 G1G1YARG1G1Y1WETG1SER1AG1YGLUARG1Y1G1Y1PHEAN1YSPROG1YGLY 20
Db 134 GGAGGAGAGGTGGCATGGCGGAGAGT--GACCGTGGTGGCTTCAATTAATTGGTGGC 190
QY 21 PrometAaspGluG1YProaspLeuaspLeuG1YProProValaspProaspGluaspSer 40
Db 191 CCTCGGAGCAGAGATCAGTCATGAC-----TCCGACAGGATTAATTC 235
QY 41 AspAsnSerAla1IeTyValG1ng1YLeuAsnAspSerValThreAspApeLeuAla 60
Db 236 GACAAACAACAACATCTTTGTCAGAGGCTGGTGAGAAATTTCAATTAATGAGTGTGGCT 295
QY 61 AspPhePheLygInCyseG1YValValLyMetAsnLyArgThrg1YgInPromet1Ie 80
Db 296 GATTACTTCAGAGAGATTGGTATTAATTAAGACAAACAAAGAAAACGGGACGCCCATGATT 355
QY 81 He1IeTyLeuaspLygInThrg1YLyseProLyseG1YAspAlaThrValSerTyrgLu 100
Db 356 AATTGTACACAGACAGGAGAACTGGCAAGCTGTAAGGAGAGGACAGGCTCTTTGAT 415
QY 101 AspProProThralaYsAlaAla-ValG1uTrp 111
Db 416 GACCCACCTTCAGCTAAAGCAGCTAATTGACTGG 449
```

```
RESULT 13
US-09-854-133-145
/ Sequence 145, Application US/09854133
/ Patent No. 6759508
/ GENERAL INFORMATION:
/ APPLICANT: Lodes, Michael J.
/ APPLICANT: Mohamath, Raodoh
/ APPLICANT: Henderson, Robert A.
/ APPLICANT: Benson, Darin R.
/ APPLICANT: Secrist, Heather
/ TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
/ TITLE OF INVENTION: THE THERAPY AND DIAGNOSIS OF LUNG CANCER
/ FILE REFERENCE: 210121.475C10
/ CURRENT APPLICATION NUMBER: US/09/854,133
/ CURRENT FILING DATE: 2001-05-11
/ NUMBER OF SEQ ID NOS: 735
/ SOFTWARE: FaastSeq for Windows Version 3.0
/ SEQ ID NO 145
/ LENGTH: 450
/ TYPE: DNA
/ ORGANISM: Homo sapien
US-09-854-133-145

Alignment Scores:
Pred. No.: 2,4e-11 Length: 450
Score: 344.00 Matches: 64
Percent Similarity: 79.468 Conservative: 25
Best Local Similarity: 57.144 Mismatches: 16
Query Match: 17.908 Indels: 7
Gaps: 2

US-10-791-017a-2_COPY_319_656 (1-338) x US-09-854-133-145 (1-450)
QY 1 G1G1YARG1G1Y1WETG1SER1AG1YGLUARG1Y1G1Y1PHEAN1YSPROG1YGLY 20
Db 134 GGAGGAGAGGTGGCATGGCGGAGAGT--GACCGTGGTGGCTTCAATTAATTGGTGGC 190
QY 21 PrometAaspGluG1YProaspLeuaspLeuG1YProProValaspProaspGluaspSer 40
Db 191 CCTCGGAGCAGAGATCAGTCATGAC-----TCCGACAGGATTAATTC 235
QY 41 AspAsnSerAla1IeTyValG1ng1YLeuAsnAspSerValThreAspApeLeuAla 60
Db 236 GACAAACAACAACATCTTTGTCAGAGGCTGGTGAGAAATTTCAATTAATGAGTGTGGCT 295
QY 61 AspPhePheLygInCyseG1YValValLyMetAsnLyArgThrg1YgInPromet1Ie 80
Db 296 GATTACTTCAGAGAGATTGGTATTAATTAAGACAAACAAAGAAAACGGGACGCCCATGATT 355
QY 81 He1IeTyLeuaspLygInThrg1YLyseProLyseG1YAspAlaThrValSerTyrgLu 100
Db 356 AATTGTACACAGACAGGAGAACTGGCAAGCTGTAAGGAGAGGACAGGCTCTTTGAT 415
QY 101 AspProProThralaYsAlaAla-ValG1uTrp 111
Db 416 GACCCACCTTCAGCTAAAGCAGCTAATTGACTGG 449

RESULT 14
US-09-249-585A-4/C
/ Sequence 48, Application US/09249585A
/ Patent No. 6417002
/ GENERAL INFORMATION:
/ APPLICANT: Hotlick, Robert
/ APPLICANT: Hotlick, Robert
/ TITLE OF INVENTION: METHOD FOR MAINTENANCE AND SELECTION OF EPISOMES
/ FILE REFERENCE: 0867/0D905
/ CURRENT APPLICATION NUMBER: US/09/249,585A
/ CURRENT FILING DATE: 1999-02-11
/ NUMBER OF SEQ ID NOS: 18
/ SOFTWARE: PatentIn version 3.0
/ SEQ ID NO 4
/ LENGTH: 1926
/ TYPE: DNA
```


ORGANISM: Epstein Barr Virus
FEATURE:
NAME/KEY: misc feature
LOCATION: (1..1926)
OTHER INFORMATION: template strand of EBNA-1 DNA
US-09-249-585A-4

Alignment Scores:

Pred. No.:	1,02e-07	Length:	1926
Score:	289.00	Matches:	131
Percent Similarity:	40.15%	Conservative:	28
Best Local Similarity:	33.08%	Mismatches:	124
Query Match:	15.04%	Indels:	114
DB:	3	Gaps:	22

US-10-791-017a-2_COPY_319_656 (1-338) x US-09-249-585A-4 (1-1926)

```
QY 1 G1YGYARGLYGLYMETGLYSERALAG1YGLUARG1YGLYPHEANLYSPROGLY--- 19
DB 1388 GAGAGAAAACCGCGACGAGAGTAGTGACCTGGGCCCCAGGTCAAGAACCGGAGAGA 1329
QY 20 G1Y-PrometAapGLUArgLysPheApleuApleuGLYProPro-----Va 34
DB 1328 GGAACCCAGTAGAC---GCCCGGGACGAGATAGCGAGCGCCCGTCAGTCCAGTGT 1272
QY 34 1AAPPProAapGLUaapSerAapSerAa1a1eTyValG1ng1YleuAaapSerVa 54
DB 1271 AGACCCGCTGAGAAACACCA---TAAGTTTATTATGACGAGG----- 1229
QY 54 1ThrlEuAapApleu1aAapPhePhe1ySg1nCySg1YVal1a1yS----- 70
DB 1228 -----GATGTCCACCTTTTATCCGAGAGATGACCTCCCGGACCGGCACCT 1179
QY 71 -----MetAen1ySArgThnG1YGLnProMet11eh1a1eTy 83
DB 1178 CTGGCCTACTACTACTGATGATGACCTGAGAGACCGGAGAAAAGAGTGGAGTGG-- 1121
QY 83 r1euaAapLyGLnThnG1YyS-----ProLySg1YAsp1a1aThyValSerTyGLUa 101
DB 1120 -CTGGAGAGGGGACCGAGAAAGTGTCTGAAAGGGGGGACCGAGAAAGTCAAGAGATGG 1062
QY 101 pProProThrA1a1yS----- 106
DB 1061 GCCGCCGAGGTATGAGAGACCTGGGCGGAGGTATGAGAGACCTGGGCGCGAGCT 1002
QY 107 -----AlaA1aValG1UTrPheAapG1YLyAapPheG1ng1ySerTy 121
DB 1001 GACGAGAGAGCTGGGCGCGAGGTGAGAGACGAGAGAGAGAGAGAGAGAGAGAG 942
QY 121 S1eulySValSer1eul1a1Arg1yS1ySProProMetAaSerMetAArg1YGLyLeuPr 141
DB 941 GGA-CAAGAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900
QY 141 oProA-Sg1nG1YArg1YMetProProProLeuAArg1YGLYProGLY-----G1YPr 159
DB 899 -----CGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 859
QY 159 oG1YGLYProGLYGLYProMetG1YArgMetG1YArgG1YGLYAspAArg1Y----- 177
DB 858 GGGAGAGAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 802
QY 178 -----G1YpHeProProAArg1YProAArg1YSerAArg1YAspRoserG1YGLY 195
DB 801 AGGACGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 742
QY 195 yAenValG1nH1a1Arg1a1Arg1YAspTrArg1nCySProAaPProGLYCySg1YAsnG1na 215
DB 741 AGCAGAGAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 662
QY 215 nPheA1aTrArgThnG1nCySAsnG1nCyS1yS1a1aProLySProGLnG1YpHeLeuPr 235
DB 681 ACGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 669
```

```
QY 235 oProProPheProProProGLYGLYAspAArg1YArgG1YGLYProGLYGLYMetAArg1 255
DB 668 -----GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 634
QY 255 YG1YArgG1YGLYLeuMetAapAArg1YGLYProGLYGLYMetPheAArg1YGLYArgG1 275
DB 633 AGGACGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 574
QY 275 YG1Y-----AspAArg1YGLYpHeAArg1YGLYArgG1YMet----- 287
DB 573 AGGACGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 514
QY 288 -----AspAArg1YGLYpHeG1YGLYArgG1YArgG1YArgG1YProGLYGLY 302
DB 513 AGGAGAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 454
QY 302 YProProGLYProLeuMetG1nMetG1YGLYArgG1YArgG1YArgG1YArgG1YPr 321
DB 453 AGGACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 406
QY 321 oG1Y-LysMetAapLySg1YGLnH1a1ArgG1nG1nUArgAArgP 335
DB 405 GGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 362

RESULT 15
US-09-130-114-2/c
/ Sequence 2, Application US/09130114
/ Patent No. 5976807
/ GENERAL INFORMATION:
/ APPLICANT: Horlick, Robert A.
/ APPLICANT: Dama, Bassem B.
/ APPLICANT: Robbins, Alan K.
/ TITLE OF INVENTION: Eukaryotic Cells Stably Expressing Genes
/ FILE REFERENCE: 0867/1D903US1
/ CURRENT APPLICATION NUMBER: US/09/130,114
/ NUMBER OF SEQ ID NOS: 36
/ SOFTWARE: FastSeq for Windows Version 3.0
/ SEQ ID NO 2
/ LENGTH: 1931
/ TYPE: DNA
/ ORGANISM: EBNA
US-09-130-114-2

Alignment Scores:
Pred. No.: 1,02e-07 Length: 1931
Score: 289.00 Matches: 131
Percent Similarity: 40.15% Conservative: 28
Best Local Similarity: 33.08% Mismatches: 124
Query Match: 15.04% Indels: 114
DB: 2 Gaps: 22

US-10-791-017a-2_COPY_319_656 (1-338) x US-09-130-114-2 (1-1931)

QY 1 G1YGYARGLYGLYMETGLYSERALAG1YGLUARG1YGLYPHEANLYSPROGLY--- 19
DB 1388 GAGAGAAAACCGCGACGAGAGTAGTGACCTGGGCCCCAGGTCAAGAACCGGAGAGA 1329
QY 20 G1Y-PrometAapGLUArgLysPheApleuApleuGLYProPro-----Va 34
DB 1328 GGAACCCAGTAGAC---GCCCGGGACGAGATAGCGAGCGCCCGTCAGTCCAGTGT 1272
QY 34 1AAPPProAapGLUaapSerAapSerAa1a1eTyValG1ng1YleuAaapSerVa 54
DB 1271 AGACCCGCTGAGAAACACCA---TAAGTTTATTATGACGAGG----- 1229
QY 54 1ThrlEuAapApleu1aAapPhePhe1ySg1nCySg1YVal1a1yS----- 70
DB 1228 -----GATGTCCACCTTTTATCCGAGAGATGACCTCCCGGACCGGCACCT 1179
QY 71 -----MetAen1ySArgThnG1YGLnProMet11eh1a1eTy 83
DB 1178 CTGGCCTACTACTACTGATGATGACCTGAGAGACCGGAGAAAAGAGTGGAGTGG-- 1121
QY 83 r1euaAapLyGLnThnG1YyS-----ProLySg1YAsp1a1aThyValSerTyGLUa 101
DB 1120 -CTGGAGAGGGGACCGAGAAAGTGTCTGAAAGGGGGGACCGAGAAAGTCAAGAGATGG 1062
QY 101 pProProThrA1a1yS----- 106
DB 1061 GCCGCCGAGGTATGAGAGACCTGGGCGGAGGTATGAGAGACCTGGGCGCGAGCT 1002
QY 107 -----AlaA1aValG1UTrPheAapG1YLyAapPheG1ng1ySerTy 121
DB 1001 GACGAGAGAGCTGGGCGCGAGGTGAGAGACGAGAGAGAGAGAGAGAGAGAGAGAG 942
QY 121 S1eulySValSer1eul1a1Arg1yS1ySProProMetAaSerMetAArg1YGLyLeuPr 141
DB 941 GGA-CAAGAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900
QY 141 oProA-Sg1nG1YArg1YMetProProProLeuAArg1YGLYProGLY-----G1YPr 159
DB 899 -----CGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 859
QY 159 oG1YGLYProGLYGLYProMetG1YArgMetG1YArgG1YGLYAspAArg1Y----- 177
DB 858 GGGAGAGAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 802
QY 178 -----G1YpHeProProAArg1YProAArg1YSerAArg1YAspRoserG1YGLY 195
DB 801 AGGACGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 742
QY 195 yAenValG1nH1a1Arg1a1Arg1YAspTrArg1nCySProAaPProGLYCySg1YAsnG1na 215
DB 741 AGCAGAGAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 662
QY 215 nPheA1aTrArgThnG1nCySAsnG1nCyS1yS1a1aProLySProGLnG1YpHeLeuPr 235
DB 681 ACGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 669
```

```
Db 1178 CTGGGCTACTACTGATGATGACCTGAGGACCCGGAGAAAAGAGTGCAGGTG--- 1121
QY rLeuAapLyvgIuThrGlyLys-----ProLyvgIyAapAlaThrValSerTygIuAs 101
   |||:::|
Db 1120 -CTGAGAGAGGGGAGCCGAGAAAGTGTGAAGGGGGGACCGAGAAAGTGCAGAGATGGG 1062
QY 101 pProPThrAlaLys----- 106
   |||
Db 1061 GTCGCGGAGAGTATGAGAGAGCTGGGGCCGAGGTGATGAGAGAGCTGGGGCCGAGAGT 1002
QY 107 -----AlaAlaValGluTrpPheAapGlyLysAapPheGlnGlySerLy 121
   |||:::|
Db 1001 GACGAGAGAGCTGGGGCCGAGGTGAGAGACGAGAGACGGGAGAGACGAGAGACGGGAGAG 942
QY 121 sLeuLyvValSerLeuAlaArgLyvLyvProPheAapSerMetArgGlyGlyLeuPr 141
   ::::
Db 941 GGA-CGAGGACGGGGAGAGCGGGGA-----GAGCGAGAGACGGGGAGAG----- 900
QY 141 oProArgGluGlyArgGlyMetProProPheAapArgGlyGlyProGly-----GlyPr 159
   |||
Db 899 ----CGGGAGAGACGAGAG-----CGGGAGAGACGGGGAGAGACGAGAGACG 859
QY 159 oGlyGlyProGlyGlyProMetGlyArgMetGlyGlyArgGlyGlyAapArgGly----- 177
   |||
Db 858 GGGAGAGACGGGGAGAGAGAGAGAGAGCG--GGAGAGACGGGGAGAGAGAGAGAGAGAGAG 802
QY 178 -----GlyPheProProArgGlyProArgGlySerArgGlyAapProSerGlyGly 195
   |||
Db 801 AGGACGGGGAGAGACGAGAGACGGGGAGAGACGGGGAGAGACGGGGAGAGACGGGGAGAG 742
QY 195 yAenValGlnHisArgLysArgTrpGlnCyvProAapProGlyCyvGlyAaenGlnAs 215
   |||
Db 741 ACAGAGACGGGGAGAGACGAGAGACGGGGAGAGACGAGAGACGGGGAGAGACGGGGAGAG 682
QY 215 nPheAlaTrpArgThrGluCyvAenGlnCyvLyvAlaProLyvProGluGlyPheLeuPr 235
   |||
Db 681 ACCGGGAGAGACGG-- 669
QY 235 oProProPheProProProGlyGlyAapArgGlyArgGlyGlyProGlyGlyMetArgGly 255
   |||
Db 668 -----GAGAGAGAGACGAGAGACGGGGAGAGACGAGAGACGGGGAGAGACGGGGAG 634
QY 255 yGlyArgGlyGlyLeuMetAapArgGlyGlyProGlyGlyMetPheArgGlyGlyArgGly 275
   |||
Db 633 AGGACGGGGAGAGACGGGGAGAGACGGGGAGAGAGAGAGAGAGACGGGGAGAGACGGGGAG 574
QY 275 yGly-----AapArgGlyGlyPheArgGlyGlyArgGlyMet----- 287
   |||
Db 573 AGGACGGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 514
QY 288 -----AapArgGlyGlyPheGlyGlyGlyArg-----ArgGlyGlyProGlyGly 302
   |||
Db 513 ACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 454
QY 302 yProProGlyProLeuMetGluGlnMetGlyGlyArg-----ArgGlyGlyArgGlyGlyPr 321
   |||
Db 453 AGGACGAGAGAG-----CGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 406
QY 321 oGly-LysMetAapLyvgIyGlnHisArgGlnGlnArgArgAap 335
   |||
Db 405 GGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 362
```

Search completed: February 21, 2005, 04:23:33
Job time : 225.958 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: February 20, 2005, 21:44:36, Search time 2535 Seconds
(without alignments)
5075.236 Million cell updates/sec

Title: US-10-791-017a-2_COPY_319_656
Perfect score: 1922
Sequence: 1 GGRGCMGSGRGRGFKPG.....GGRGKDKGHRGRDRRY 338

Scoring table:
BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 34239544 seqs, 19032134700 residues
Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODE=frame+.p2n.model -DEV=xip
-Q/cg2.1/USPTO.spool.p/US10791017/runat.17022005.125807.22086/app.query.fasta_1.1358
-DB=EST -QFMT=faa2ap -SUFFIX=p2n.rst -MINMATCH=0.1 -LOOPT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blomsum62 -TRANS=human40.cdd -LIST=45
-LOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pco -NOR=ext -HEATSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10791017@cgn.1.1.6628@runat.17022005.125807.22086 -NCPU=6 -ICPU=3
-NO_WMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -NARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :
EST:
1: gb_esc1:*
2: gb_esc2:*
3: gb_hic:*
4: gb_esc3:*
5: gb_esc4:*
6: gb_esc5:*
7: gb_esc6:*
8: gb_gsa1:*
9: gb_gsa2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1922	100.0	2073	CR593724	full-length
2	1922	100.0	2083	CR593724	full-length
3	1922	100.0	2103	CR619493	full-length
4	1922	100.0	2157	CR604952	full-length
5	1922	100.0	2314	CR608375	full-length
6	1907.5	99.2	2070	CR602561	full-length
7	1907.5	99.2	2093	CR594868	full-length
8	1907.5	99.2	2126	CR610888	full-length
9	1907.5	99.2	2159	CR625247	full-length

10	1888.5	98.3	2107	3	AK019460	AK019460	Mus muscu
11	1874	97.5	1379	3	AK014366	AK014366	Mus muscu
12	1871.5	97.4	2237	3	AK034755	AK034755	Mus muscu
13	1744.5	90.8	2376	3	CR604639	CR604639	full-length
14	1719.5	89.5	2269	3	AK049743	AK049743	Mus muscu
15	1675.5	87.2	4729	3	HSMB04802	AL833489	Hom sapi
16	1608.5	83.7	1787	3	CR618632	CR618632	full-length
17	1537.5	80.0	1895	5	EX408406	EX408406	EX408406
18	1537	80.0	1020	5	EX387336	EX387336	EX387336
19	1496.5	77.9	929	5	BQ881305	BQ881305	AGENCOURT
20	1495.5	77.8	1068	5	EX449773	EX449773	EX449773
21	1495	77.8	843	7	CN162640	CN162640	952042 MA
22	1482	77.1	897	7	BY713354	BY713354	BY713354
23	1470.5	76.5	1057	5	EX353312	EX353312	EX353312
24	1464	76.2	962	5	EX407870	EX407870	EX407870
25	1456.5	75.8	867	4	BI647600	BI647600	603278181
26	1450.5	75.5	940	5	BQ927930	BQ927930	AGENCOURT
27	1435	74.7	1013	5	EX465099	EX465099	EX465099
28	1435	74.7	1048	5	EX381313	EX381313	EX381313
29	1429	74.3	838	4	BG574437	BG574437	602596384
30	1428	74.3	1028	5	EX399069	EX399069	EX399069
31	1406	73.2	778	4	BG756215	BG756215	602713525
32	1403	73.0	1055	5	EX428311	EX428311	EX428311
33	1400.5	72.9	1026	5	BQ608080	BQ608080	AGENCOURT
34	1390	72.3	1022	5	EX407825	EX407825	EX407825
35	1386	72.1	792	5	EX368129	EX368129	EX368129
36	1377.5	71.7	837	4	BG751179	BG751179	60272679
37	1377	71.6	744	7	CR940579	CR940579	4063812 B
38	1373	71.4	925	5	EX333942	EX333942	EX333942
39	1372	71.4	933	5	EX387305	EX387305	EX387305
40	1368.5	71.2	2632	3	BC043904	BC043904	Xenopus 1
41	1368	71.2	901	2	BE562296	BE562296	601344870
42	1363	70.9	715	5	BQ042484	BQ042484	UI-M-EMO-
43	1362	70.8	737	5	EX919060	EX919060	EX919060
44	1361	70.9	828	5	BQ671992	BQ671992	AGENCOURT
45	1347	70.1	787	2	BE741903	BE741903	601594075

ALIGNMENTS

RESULT 1
CR593724
LOCUS
DEFINITION
full-length cDNA clone CS0D1011Y05 of Placenta Cot 25-normalized
of Homo sapiens (human).
ACCESSION
CR593724.1 GI:50474531
VERSION
CR593724.1
KEYWORDS
HTC; CNSIT_CDNA.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 2073)
Li, W.B., Gruber, C., Jessee, J. and Polyes, D.
Full-length cDNA libraries and normalization
Unpublished
JOURNAL
Contact: Feng Liang Email: fliang@life.techn.com URL:
http://fulllength.invitrogen.com/ invitrogen Corporation 1600
Faraday Avenue
2 (bases 1 to 2073)
REFERENCE
Genoscope.
Direct Submission
Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail: seqref@genoscope.cns.fr
- Web: www.genoscope.cns.fr)
COMMENT
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by Life Technologies, a
division of Invitrogen.
FEATURES
source
1..2073
/organism="Homo sapiens"

ORIGIN

/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0D1011Y005"
/issue_type="Placenta Cot 25-normalized"
/plasmid="pCMVSPORT_6"

Alignment Scores:

Pred. No.: 4,11e-109 Length: 2073
Score: 1922.00 Matches: 338
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
Gaps: 0

US-10-791-017a-2_copy_319_656 (1-338) x CR593724 (1-2073)

QY 1 G1YGLYARGLYGLYMETGLYSERIALAGLYGLUARGLYGLYPHEANLYSPROGLYGLY 20
DB 906 GAGAGACGCGGTGGATGGGCAAGCGCTGGAGAGCGAGTGGCTTCAATTAAGCCTGGTGA 965
QY 21 PROMETASPNGLUGLYPROASPLEUASPLEUGLYPROBROVALASPROMASPLUASPSER 40
DB 966 CCCATGGATGAGAGACCAAGATCTTGATCTAGGCCCACTGTGATGATCCAGATGAGACTCT 1025
QY 41 ASPASNSERIALLEYRYVALGINGLYLEUASNASPSERVALTHYLEUASPLEUALA 60
DB 1026 GACAAACGTCGCAATTTATGTACAGAGATTAAATGACGTGTGACTTGAATGATCTGGCA 1085
QY 61 ASPHAPHELYSGINCYSGLYVALVALYLMERASNYLSAQTTHRGLYGLINPROMETILE 80
DB 1086 GACTTCTTTAAGCAGTGTGGGTTGTTAAGTGAACAGAGAACTGGGCAACCATATC 1145
QY 81 H1E1ETYLEUASPLYSGIUTHRGILYSPROLYSGIYASPA1ATHRVALSERTYRGILU 100
DB 1146 CACATCTACCTGACCAAGAAACAGAAAGCCCAAGAGCCATGACAGTGTCTATGAA 1205
QY 101 ASPPROBOTHRALALYVALALVALAGLUTRPHASPGIYLYSASPHEGLINGLYSER 120
DB 1206 GACCCACCCCACTGGCAAGGCTGGCGGTGAATGATTTATGGGAAAGATTTTCAAGGAGC 1265
QY 121 LYSLEUASVALSERLEUALAARGLYLYSPROBROMECASNSERMETARGLYGLYLEU 140
DB 1266 AATCTTAAAGTCTCCCTTCTCGAAGAGCCTTCAATGAACATATGCGGGGTGCTG 1325
QY 141 PROBROARGIUGLYARGIYMETPROBROBROLEUARGIYGLYPROGLYGLYPROGLY 160
DB 1326 CCAACCCGTAAGGAGCAAGAGCATCCACCACTCGGTGAGGTCCAGAGAGGCCAGGA 1385
QY 161 GLYPROGLYGLYPROMETGLYARGMETGLYGLYARGGLYGLYASPAARGLYGLYPHEPRO 180
DB 1386 GGTCTGGGGAGCCCAATGGGTGCGCATGGAGCGCGTGGAGAGATTAAGAGGCTTCCCT 1445
QY 181 PROARGIYPROARGIYSPERARGIYASNPBROSERGLYGLYAAVNAIGH1SARG 200
DB 1446 CCAAGAGAGACCCCGGGGTTCGAGAGGAAACCTCTGGAGAGAGAAAGTTCACACCGGA 1505
QY 201 ALAAGLYASPTRGINCYSPROASNPARGIYCYSGLYASNGINANPHEALATPRARGTHR 220
DB 1506 GCTGGAAGCTGGCAGTGTCCCAATCCGGGTGTGGAACCAAGAACTTCCGCTGAGAAACA 1565
QY 221 GLUCYASANGINCYSLYVALAASPROLYSPROGLUGLYPHELEUPROPROPHASPROPRO 240
DB 1566 GAGTGAACAACAGTGAAGGCCCCCAAGGCTTCTCCGCGCAACCTTTTCCGCGCC 1625
QY 241 PROGLYGLYASPAARGIYARGIYGLYPROGLYGLYMETARGIYGLYARGIYGLYLEU 260
DB 1626 CCGGATGTGATCGTGGAGAGGTGGCCCTGATGCAAGCGGAGAGAGAGGTGGCTC 1685
QY 261 METASPAARGIYGLYPROGLYGLYMETPHEARGIYGLYARGIYGLYGLYASPAARGIYGLY 280
DB 1686 ATGATCTGTGTGTCCGCGGTGGATGTTCAGAGGTGGCCGTGTGAGAGAGAGAGGTGGC 1745

QY 281 PHEARGIYGLYARGIYMETASPAARGIYGLYPHEGLYGLYARGIYGLYARGIYGLYPRO 300
DB 1746 TTCGGTGTGGTGGCGGCGCATGAGACGAGTGGCTTTGGTGAAGAGAGAGGTGGCCCT 1805
QY 301 G1YGLYPROGLYPROLEUWETGLUGINMETGLYGLYARGIYGLYGLYARGIYGLY 320
DB 1806 GGGGGGGCCCCCTGACCTTTGATGAGAACAGTGGAGAGAGAGAGAGAGAGAGAGAGAG 1865

QY 321 PROGLYLYMETASPLYGLYGLYH1SARGIINGLYARGIYARGIYASPAARGPROTYR 338
DB 1866 CCGGAAAAATGATTAAGCCAGACACGTCAGAGCGGAGAGATGGCCCTTAC 1919

RESULT 2
CR608666
LOCUS
DEFINITION
full-length cDNA clone CS0D004Y123 of B cells (Ramos cell line) of
Homo sapiens (human).

ACCESSION
CR608666
VERSION
CR608666.1 GI:50489473
KEYWORDS
HTC; CNST; cDNA.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
1 (bases 1 to 2083)
AUTHORS
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
TITLE
Full-length cDNA libraries and normalization
JOURNAL
Unpublished
CONTACT : Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Faraday Avenue
2 (bases 1 to 2083)

REFERENCE
AUTHORS
Genoscope.
TITLE
Direct Submission
JOURNAL
Submitted (20-JUN-2004) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
COMMENT
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by Life Technologies, a
division of Invitrogen.
FEATURES
Location/Qualifiers
1..2083
source
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0D004Y123"
/issue_type="B cells (Ramos cell line)"
/plasmid="pCMVSPORT_6"

ORIGIN

Alignment Scores:

Pred. No.: 4,14e-109 Length: 2083
Score: 1922.00 Matches: 338
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
Gaps: 0

US-10-791-017a-2_copy_319_656 (1-338) x CR608666 (1-2083)

QY 1 G1YGLYARGLYGLYMETGLYSERIALAGLYGLUARGLYGLYPHEANLYSPROGLYGLY 20
DB 906 GAGAGACGCGGTGGATGGGCAAGCGCTGGAGAGCGAGTGGCTTCAATTAAGCCTGGTGA 965
QY 21 PROMETASPNGLUGLYPROASPLEUASPLEUGLYPROBROVALASPROMASPLUASPSER 40
DB 966 CCCATGGATGAGAGACCAAGATCTTGATCTAGGCCCACTGTGATGATCCAGATGAGACTCT 1025
QY 41 ASPASNSERIALLEYRYVALGINGLYLEUASNASPSERVALTHYLEUASPLEUALA 60
DB 1026 GACAAACGTCGCAATTTATGTACAGAGATTAAATGACAGTGTGACTTGAATGATCTGGCA 1085

QY 61 AspPhePheIysGInCysGlyValValIysMetAsnIysArgThrGlyGlnPrometIle 80
 1086 GACTTCTTTAAGCAGTGTGGGTTGTTAAGTATGAACAGAGACTGGGCAACCATATATC 1145
 QY 81 HisIleTyrLeuAspIysGlnThrGlyLysProIysGlyAspAlaThrValSerTyrGlu 100
 1146 CACATCTACCTGGACAGAAACAGAAAGCCCAAGGCGATGCGACAGTGTCTATGAA 1205
 QY 101 AspProProThrAlaIysAlaIaValGluTTPheAspGlyLysAspPheGlnIysSer 120
 1206 GACCCACCCCACTGCGAAGGCTGCGCTGGAATGTGTTATGGGAAAGATTTTCAAGGAGC 1265
 QY 121 LysLeuLysValSerIleuAlaArgLysLysProMetAsnSerMetArgGlyLysLeu 140
 1266 AATCTTAAGTCTCTCTGCTGGAGAGAGCTTCATGAACAGTATGCGGGGTGCTG 1325
 QY 141 ProProArgGlnGlyArgGlyMetProProProLeuArgGlyLysProGlyLysProGly 160
 1326 CCAACCCCGTGAAGGAGAGGAGATGCCACACCACTCGTGGAGGTCCAGAGAGGCCAGGA 1385
 QY 161 GlyProGlyLysProMetGlyArgMetGlyLysArgGlyLysAspArgGlyLysPhePro 180
 1386 GGTCTGGGGAGACCATGGGTCCGATGCGAGGCGCTGAGAGAGATAGAGAGGCTTCCCT 1445
 QY 181 ProArgGlyProArgGlySerArgGlyLysAsnProSerGlyLysGlyAsnValGlnHisArg 200
 1446 CCAAGAGAGACCCCGGGGTCTCCAGAGGAGACCTCTGAGAGAGAAACGTCAGACACGA 1505
 QY 201 AlaGlyAspTTPGInCysProAsnProGlyCysGlyLysGlnAsnPheAlaTTPArgThr 220
 1506 GCTGGAGACTGGCAGTGTCCCATCCGGGTGTGGAAACAGAACTTCGCTGGAGAGCA 1565
 QY 221 GlnCysAsnGlnCysLysAlaProLysProGlnGlyLysPheLeuProProPhePro 240
 1566 GAGTGCACACGATGAAGGCCCAAGACCTGAAGGCTTCTCCGCGCACCTTCCGCC 1625
 QY 241 ProGlyLysAspArgGlyArgGlyLysProGlyLysMetArgGlyLysArgGlyLysLeu 260
 1626 CCGGGGTGTATGTGGCAGAGGTGGCTGTGTGGCATGTGGGAGAGAGAGGTGGCTTC 1685
 QY 261 MetAspArgGlyLysGlyProGlyLysMetPheArgGlyLysArgGlyLysAspArgGlyLys 280
 1686 ATGATCGTGTGTGTCCCGGTGATTTTCAAGAGTGTGGCTGTGGAGCAGAGGTGSC 1745
 QY 281 PheArgGlyLysArgGlyMetAspArgGlyLysPheGlyLysGlyLysArgGlyLysPro 300
 1746 TTCCTGTGTGGCGGGCATGACCGAGGTGTGGTGTGGAGAGAGAGAGGTGGCTTC 1805
 QY 301 GlyLysProProGlyProLeuMetGlnGlnMetGlyLysArgArgGlyLysArgGlyLys 320
 1806 GGGGGGCCCCCTTGAACCTTGTGAAGCAAGATGGAGAGAGAGAGAGAGAGAGAGAGAG 1865
 QY 321 ProGlyLysMetAspLysGlyLysHisArgGlnGlnArgArgAspArgProTyr 338
 1866 CCTGGAAATATGATTAAGGCGAGACCGTCAAGAGCGCAGAGATGGCCCTTAC 1919
 Db
 RESULT 3
 CR619493 2103 bp mRNA linear HTC 21-JUL-2004
 LOCUS CR619493
 DEFINITION Full-length cDNA clone CS001076Y021 of Placenta Cot 25-normalized
 of Homo sapiens (human).
 ACCESSION CR619493
 VERSION CR619493.1 GI:50500300
 KEYWORDS HTC; CNSLT cDNA.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE
 AUTHORS Li, W.-B., Gruber, C., Jessee, J., and Polayes, D.
 TITLE Full-length cDNA libraries and normalization
 JOURNAL unpublished

REMARK Contact : Peng Liang Email : liang@life.techn.com URL :
 http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
 Redwood Avenue
 Redwood City, CA 94063
 REFERENCE 2 (bases 1 to 2103)
 AUTHORS Genoscope.
 TITLE Direct Submission
 JOURNAL Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :
 BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
 - Web : www.genoscope.cns.fr)
 COMMENT 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
 end enriched, double-strand cDNA was digested with Not I and cloned
 into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
 was normalized. Library was constructed by Life Technologies, a
 division of Invitrogen.
 FEATURES
 source
 1. 2103
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="CS001076Y021"
 /tissue_type="Placenta Cot 25-normalized"
 /plasmid="pCMVSPORT_6"
 ORIGIN
 Alignment Scores:
 Pred. No.: 4.18e-109 Length: 2103
 Score: 1922.00 Matches: 338
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 3 Gaps: 0
 US-10-791-017A-2_COPY_319_656 (1-338) x CR619493 (1-2103)
 QY 1 GlyLysArgGlyLysLysMetGlySerAlaGlyLysArgGlyLysPheAsnLysProGlyLys 20
 943 GAGAGACCGGTGTGATGGCAGACCTGAGAGCGAGGTGCTTCAATAGCCTGTGGGA 1002
 QY 21 PrometAspGlnGlyProAspLeuAspLeuGlyProProValAspProAspGluAspSer 40
 1003 CCCATGATGAAGACCAAGATCTTGAATCTAGGCGCACCTGTATGATCCAGATGAAGACTCT 1062
 QY 41 AspAsnSerAlaIleTyrValGlnGlnLysMetAsnLysArgThrLeuAspAspLeuAla 60
 1063 GACCAACAGTGCATTTATGTACAAAGATTAAAGAGATGTGACTAGATGATCTGGCA 1122
 QY 61 AspPhePheIysGInCysGlyValValIysMetAsnIysArgThrGlyGlnPrometIle 80
 1123 GACTTCTTTAAGCAGTGTGGGTTGTTAAGTATGAACAGAGACTGGGCAACCATATATC 1182
 QY 81 HisIleTyrLeuAspIysGlnThrGlyLysProIysGlyAspAlaThrValSerTyrGlu 100
 1183 CACATCTACCTGGACAGAAACAGAAAGCCCAAGGCGATGCGACAGTGTCTATGAA 1242
 QY 101 AspProProThrAlaIysAlaIaValGluTTPheAspGlyLysAspPheGlnIysSer 120
 1243 GACCCACCCCACTGCGAAGGCTGCGCTGGAATGTGTTATGGGAAAGATTTTCAAGGAGC 1302
 QY 121 LysLeuLysValSerIleuAlaArgLysLysProMetAsnSerMetArgGlyLysLeu 140
 1303 AATCTTAAGTCTCTCTGCTGGAGAGAGCTTCATGAACAGTATGCGGGGTGCTG 1362
 QY 141 ProProArgGlnGlyArgGlyMetProProProLeuArgGlyLysProGlyLysProGly 160
 1363 CCAACCCCGTGAAGGAGAGGAGATGCCACACCACTCGTGGAGGTCCAGAGAGGCCAGGA 1422
 QY 161 GlyProGlyLysProMetGlyArgMetGlyLysArgGlyLysAspArgGlyLysPhePro 180
 1423 GGTCTGGGGAGACCCATGGGTGCGATGGAGGCGCTGTGGAGAGATAGAGAGGCTTCCCT 1482
 QY 181 ProArgGlyProArgGlySerArgGlyLysAsnProSerGlyLysGlyLysValGlnHisArg 200
 1483 CCAAGAGACCCCGGGGTCTCCGAGAGAACCTCTCTGTGGAGAGAGAGAGAGAGAGAGAG 1542

RESULT 5
CR608375
LOCUS
DEFINITION
full-length cDNA clone CS0D1039Y24 of Placenta Cot 25-normalized
of Homo sapiens (human).
ACCESSION
CR608375
VERSION
CR608375.1 GI:50489182
KEYWORDS
HTC; CNSLT_CDNA.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 2314)
Li, W.B., Gruber, C., Jesssee, J. and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished
Contact : Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/Invitrogen Corporation 1600
Faraday Avenue
2 (bases 1 to 2314)
Genoscope.
REFERENCE
AUTHORS
JOURNAL
TITLE
COMMENT
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by Life Technologies, a
division of Invitrogen.
FEATURES
source
Location/Qualifiers
1..2314
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0D1039Y24"
/tissue_type="Placenta Cot 25-normalized"
/plasmid="pCMVSPORT_6"
ORIGIN
Alignment Scores:
Pred. No.: 4.63e-109 Length: 2314
Score: 1922.00 Matches: 338
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 3 Gaps: 0
US-10-791-017a-2_COPY_319_656 (1-338) x CR608375 (1-2314)
QY 1 GlyIyAArgIyGlyMetGlySerIaGlyIyAArgIyGlyIyPheAAsnlyPProGlyIy 20
Db 995 GGAAGACGCGCGTGAATGGGCAAGCGCTGAGAGAGAGGTGGCTTCATATAGCCTGGTGA 1054
QY 21 ProMetAAspGluGlyProAAspLeuAAspLeuGlyProProValAAspProAAspGluAAspSer 40
Db 1055 CCATGATGATGAAGACCAAGATCTGTGATCTGGGCCACCTGTAGATCCAGATGAAGACTCT 1114
QY 41 AAspAAspSerIaIyIyValIyGlnIyLeuAAspSerValThreIyAAspAAspLeuA 60
Db 1115 GACAAACAGTCAATTTATGTACAGATTAATGACAGTGTGACTGTAGATGATCTGGCA 1174
QY 61 AAspPhePheIyGlnIyCysGlyValIyValIyMetAAsnlyAAspGlyIyGlnIyPheIy 80
Db 1175 GACTTCTTAAAGAGGTGGGTGGTTGTTAATGATGAACAGGAAGCTGGCAACCAAGATC 1234
QY 81 HleIyIyIyLeuAAspIyGlyIy 100
Db 1235 CACATCTACTGACAGCAAGAAAGAAAGCCCAAGGCGCATGCCACAGTGTCTTATGA 1294
QY 101 AAspProThreIy 120

Db 1295 GACCCACCACTGCCAAGCTGCCGTGAATGTTGATGGAAAGATTTTCAAGGAGC 1354
QY 121 LysLeuIyValIySerIyLeuAAspIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIy 140
Db 1355 AAACCTTAAGTCTCCCTGCTCGGAAGAAAGCTTCATATGAAGATATGCGGGGTGCTG 1414
QY 141 ProProAAspGluIy 160
Db 1415 CCAACCCGTGAAGGAGAGGAGATGACCAACCACTCGTGAAGGTCCAGAGGCCAGGA 1474
QY 161 GlyIyProGlyIy 180
Db 1475 GGTCTGGGAGACCCATGAGTCCGATGAGAGCCGTGAGAGAGATGAGAGAGCTTCCCT 1534
QY 181 ProAArgIyProAArgIyIySerAArgIyAAspProSerGlyIyIyIyIyIyIyIyIyIy 200
Db 1535 CCAAGAGAGACCCGGGGGTCCCGAGGAGACCCCTCTGAGAGAGAAACGTCCAGACGA 1594
QY 201 AlAGIyAAspTTPGlnIy 220
Db 1595 GCTGAGAGACTGGAGAGTGTCCCAATCCGGGTGGAGAAACCAAGACTTCGCTGGAGACA 1654
QY 221 GluCyAAsnIyCysIyValIyProIy 240
Db 1655 GAGTGAACCAAGTGTAAAGGCCCAAGAGCTTCTTCCGACACCTTCCGACC 1714
QY 241 ProGlyIyIyAAspAArgIy 260
Db 1715 CCGGAGTGTATGCTGGCAGAGGTGGCCCTGTGGATGCGGGAGAGAAAGAGGTGGCTC 1774
QY 261 MetAAspAArgIy 280
Db 1775 ATGATCTGTGTGTGTCCCGTGAATGTTCAAGGTGGCCGTGTGTGAACAAGAGTGGC 1834
QY 281 PheAArgIyIyIyAArgIy 300
Db 1835 TTCCGTGTGTGTGGCGGAGATGACCAAGGTGGCTTTGTGTGAAGAAACAGAGTGGCTC 1884
QY 301 GlyIyIyProProGlyIy 320
Db 1895 GGGGGGCCCCCTGACCTTGTATGAGAACAGATGGAGAAAGAGAGAGACGTGAGGA 1954
QY 321 ProGlyIyIyMetAAspIyGlyIy 338
Db 1955 CCTGAAAAATGATTAAGCGAGACCGTCAGAGCGCAGAGATGGCCCTAC 2008
RESULT 6
CR602561
LOCUS
DEFINITION
full-length cDNA clone CS0D106BYN23 of Placenta Cot 25-normalized
of Homo sapiens (human).
ACCESSION
CR602561
VERSION
CR602561.1 GI:50483368
KEYWORDS
HTC; CNSLT_CDNA.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 2070)
Li, W.B., Gruber, C., Jesssee, J. and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished
Contact : Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/Invitrogen Corporation 1600
Faraday Avenue
2 (bases 1 to 2070)
Genoscope.
REFERENCE
AUTHORS
JOURNAL
TITLE
COMMENT
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned

into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies, a division of Invitrogen.

FEATURES

source

1. .2070
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS01064YE15"
/issue_type="Placenta Cot 25-normalized"
/plasmid="pCMVSPORT_6"

ORIGIN

Alignment Scores:

Pred. No.: 3.24e-108 Length: 2070
Score: 1907.50 Matches: 337
Percent Similarity: 99.70% Conservative: 0
Best Local Similarity: 99.70% Mismatches: 0
Query Match: 99.25% Indels: 1
DB: 3 Gaps: 1

US-10-791-017a-2_COPY_319_656 (1-338) x CR602561 (1-2070)

1 G1G1YARGLYGLYMETGLYSERAG1AG1G1UARG1YGL1YPhaenylsProGlyGly 20
943 GAGAGACGCGGTGAAATGGC---GCTGAGAGCGAGGTGGCTTCATAGCCTGGTGA 999
21 ProMetAspGluGlyProAspLeuAspLeuGlyProProValAspProAspGluAspSer 40
1000 CCGATGATGAAAGACAGATCTTGATCTAGTCCACCTTAGATCCAGATGAAGACTCT 1059
41 AspAspSerAlaIleTyrValGlnGlyLeuAspSerValThrLeuAspAspLeuAla 60
1060 GACACACGTGCAATTATGTCAGAGATTAATATACAGTGTGACTTATGATGATCTGGCA 1119
61 AspPhePheLysGlnCysGlyValValLysMetAsnLysArgThrGlyGlnProMetIle 80
1120 GACTTCTTTAAGCAGTGTGGGTGTTTAAATGTAAGCAAGAGAACTGGGCAACCATGATC 1179
81 HisIleTyrLeuAspLysGluThrGlyLysProLysGlyAspAlaThrValSerTyrGlu 100
1180 CACATCTTAAAGTCTCCCTTCTCGGAAGAGCCTCCATGAAACAGTATGCGGGGTGCTG 1239
101 AspProProThrAlaLysAlaAlaValGluThrPheAspGlyLysAspPheGlnGlySer 120
1240 GACCCACCACTGCAAGGCTGCGGTGAAATGTTTAAATGTAAGTTTCAAGGAGC 1299
121 LysLeuLysValSerLeuAlaArgLysLysProMetAsnSerMetArgGlyGlyLeu 140
1300 AAACTTAAAGTCTCCCTTCTCGGAAGAGCCTCCATGAAACAGTATGCGGGGTGCTG 1359
141 ProProArgGluGlyArgGlyMetProProProLeuArgGlyGlyProGlyGlyProGly 160
1360 CCAACCCGCGTGAAGGCAAGGAGCATCCACACATCCGTGAGGTCCAGAGGCGCCAGGA 1419
161 GlyProGlyGlyProMetGlyLysMetGlyLysArgGlyGlyAspArgGlyGlyPhePro 180
1420 GGTCTGGGGAGCCAGTGGTGGATGGAGGCGGTGAGAGATGAGAGGCTTCCCT 1479
181 ProArgGlyProArgGlyLysSerArgGlyLysProSerGlyGlyGlyAsnValGlnHisArg 200
1480 CCAAGAGGACCCCGGGGTTCGCGAGGAGAACCTCTGAGAGAGAAAGCTCCACACAGGA 1539
201 AlaGlyAspTyrGlnCysProAsnProGlyCysGlyAsnGlnAsnPheAlaTyrArgThr 220
1540 GCTGAGAGCTGGAGTGTCCCAATCCGGGTGTGGAAACCAAGACTTCCGTGAGAGGA 1599
221 GlnCysAsnGlnCysLysValAlaProLysAspProGlyGlyPheLeuProProProProPro 240
1600 GAGTGAACCAAGTGAAGGCCCCAAAGCTGAAGGCTTCTCCGCAACCTTTCGCGCC 1659
241 ProGlyLysAspArgGlyArgGlyGlyProGlyGlyMetArgGlyGlyArgGlyGlyLeu 260

1660 CCGGAGTGTGATCTGAGCAGAGTGGCCCTGTGGCATCGGGGAGGAAGAGTGGCTTC 1719
261 MetAspArgGlyGlyProGlyGlyMetPheArgGlyGlyArgGlyGlyAspArgGlyGly 280
1720 ATGATCTGTGTGTCCTCGGTGAATGTTCAGAGGTGGCGGTGGAGACAGAGGTGGC 1779
281 PheArgGlyGlyArgGlyLysMetAspArgGlyGlyPheGlyGlyGlyArgArgGlyGlyPro 300
1780 TTCGGTGTGGCGGGGAGTGAACGAGGTGGCTTGGTGAAGAAAGACAGAGTGGCCCT 1839
301 GlyGlyProProGlyProLeuMetGluGlnMetGlyGlyArgArgGlyGlyArgGlyGly 320
1840 GCGGGGCGCCCTGAGCTTGTGATGAAACAGATGGGAGGAAGAGAGAGACGATGGAGGA 1899
321 ProGlyLysMetAspLysGlyGlyHisArgGlnGlyArgArgAspProTyr 338
1900 CCGGAAAAATGATTAAGGCGAGCACCTGTCAGAGCGGAGATCGCCCTTAC 1953
RESULT 7
LOCUS CR594868
DEFINITION CR594868 2093 bp mRNA linear HTC 21-JUL-2004
full-length cDNA clone CS01064YE15 of Placenta Cot 25-normalized
of Homo sapiens (human).
ACCESSION CR594868
VERSION CR594868.1 GI:50475675
KEYWORDS HTC; CNSLT cDNA.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 2093)
AUTHORS Li, W.B., Gruber, C., Jessee, J., and Polayes, D.
JOURNAL Full-length cDNA libraries and normalization
REMARK Unpublished
Contact: Feng Liang Email: fliang@lifetech.com URL: http://fulllength.invitrogen.com/ Invitrogen Corporation 1600 Faraday Avenue
2 (bases 1 to 2093)
REFERENCE Genoscope.
AUTHORS Direct Submission
JOURNAL Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail: seqref@genoscope.cns.fr)
COMMENT - Web : www.genoscope.cns.fr)
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies, a division of Invitrogen.

FEATURES

source

1. .2093
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS01064YE15"
/issue_type="Placenta Cot 25-normalized"
/plasmid="pCMVSPORT_6"

ORIGIN

Alignment Scores:

Pred. No.: 3.28e-108 Length: 2093
Score: 1907.50 Matches: 337
Percent Similarity: 99.70% Conservative: 0
Best Local Similarity: 99.70% Mismatches: 0
Query Match: 99.25% Indels: 1
DB: 3 Gaps: 1

US-10-791-017a-2_COPY_319_656 (1-338) x CR594868 (1-2093)

1 G1G1YARGLYGLYMETGLYSERAG1AG1G1UARG1YGL1YPhaenylsProGlyGly 20
943 GAGAGACGCGGTGAAATGGC---GCTGAGAGCGAGGTGGCTTCATAGCCTGGTGA 999
21 ProMetAspGluGlyProAspLeuAspLeuGlyProProValAspProAspGluAspSer 40

Db 1000 CCCATGATGACAGGACCAAGCTTCTGATCTGAGCCCACTTGAATGATGAAAGCTCT 1059

Qy 41 AapAaSerAla1leYrYalGInGlyLeuAaAapSerValThrLeuAaApleuAla 60

Db 1060 GACAAACAGTCAATTTATGTACAAAGATTAATGACAGTGTGACTGATGATGATCTGGCA 1119

Qy 61 AapPhePheLeuGInGlyValValValValValValValValValValValValValVal 80

Db 1120 GACTTCTTTAAGCAGTGTGGGTTGTTAAGATGAACAGAAAGAACTGGGCAACCATGATC 1179

Qy 81 Hs1e1eYrLeuAaApleuGInGlyValValValValValValValValValValValVal 100

Db 1180 CACATCTACCTGACAAAGAAACAGAAAGCCCAAGGCAATGCCACAGTCTCTATGAA 1239

Qy 101 AapProProThra1aYalValAlaValAlaValAlaValAlaValAlaValAlaValAla 120

Db 1240 GACCCACCCCACTGCTCCAGGCTGCGTGGAAATGGTTGATGGAAAGATTTTCAAGGAGC 1299

Qy 121 LysLeuLysValSerLeuAlaArgLysLysProProMetLysSerMetLysArgLysLys 140

Db 1300 AAACCTTAAGTCTCCCTGCTCGAAGAAAGCTTCAATGAACAGTATGCGGGTGGCTG 1359

Qy 141 ProProArgGluGlyValArgGlyMetProProProLeuAaArgGlyValProGlyVal 160

Db 1360 CCAACCCCGTGAAGGAGGAGGCAATGCCACCACTCCGTGAGAGTCCAGAGGCGCAAGA 1419

Qy 161 GlyProGlyValProMetGlyValArgMetGlyValArgGlyValArgGlyValArgGly 180

Db 1420 GGTCTGGGGGACCAATGGGTGCAATGGAGGCGGTGGAGAGATGAGAGAGGCTTCCCT 1479

Qy 181 ProArgGlyProArgGlySerArgGlyValAaProSerGlyValGlyValValGlnHsArg 200

Db 1480 CCAAGAGAGACCCCGGGGTTCCCAAGAGAACCTCTGAGAGAGAAACGTCAGACCGCA 1539

Qy 201 AlaGlyAaPTrpGInGlyProAaProArgGlyValGlyValAaAaPheAlaTrpArgGly 220

Db 1540 GCTGGAGACTGGCAAGTGTCCCAATCCGGGTTGTGAAACCAAGAACTTGGCTGAGAA 1599

Qy 221 GlnCyAaAaGInGlyValAaAaProLysProGluGlyPheLeuProProProPheProPro 240

Db 1600 GATGTCACCAAGTGTAAAGGCCCAAGGCTGAAAGGCTTCTCCGCAACCTTTCCGGCC 1659

Qy 241 ProGlyValAaAaArgGlyValArgGlyValArgGlyValArgGlyValArgGlyVal 260

Db 1660 CCGGAGTGTATGTGTGAGAGAGTGTGCTGCTGTGATGCGGAGAGAGAGAGTGTGCTC 1719

Qy 261 MetAaAaArgGlyValProGlyValMetPheArgGlyValArgGlyValArgGlyVal 280

Db 1720 ATGATCTGTGTGTGCTCCGGTGAATGTTCAAGAGTGTGCTGTGTGAGACAGAGTGTG 1779

Qy 281 PheArgGlyValArgGlyValMetAaAaArgGlyValPheGlyValArgGlyValArgGly 300

Db 1780 TTCCTGTGTGTGCGGGGCAATGAGCCGAGTGTGCTTTGTGTGAGAAAGAGAGTGTG 1839

Qy 301 GlyGlyProProGlyProLeuMetGlnMetGlyValArgGlyValArgGlyValArgGly 320

Db 1840 GGGGGGCGCCCTGTGACCTTTGATGAGAAACAGATGTGAGAAAGAGAGAGAGAGAGAG 1899

Qy 321 ProGlyLysMetAaAaArgGlyValHsArgGlnGlnAaArgAaAaArgProGly 338

Db 1900 CCTGGAAATGATTAAGGCGAGCACCGTCAAGAGGCGAGAGATGCGCCCTTAC 1953

RESULT 8

CR610888 2126 bp mRNA linear HMC 21-JUL-2004

LOCUS CR610888

DEFINITION full-length cDNA clone CS0DF037Y107 of Fetal brain of Homo sapiens (human).

ACCESSION CR610888

VERSION CR610888.1 GI:50491695

KEYWORDS HMC; cNSLT; cDNA.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 2126)

AUTHORS Li, W.B., Gruber, C., Jessee, J., and Polayes, D.

TITLE Full-length cDNA libraries and normalization

JOURNAL Unpublished

REMARK Contact: Feng Liang Email: fliang@lifetech.com URL: http://fulllength.invitrogen.com/ Invitrogen Corporation 1600 Faraday Avenue

REFERENCE 2 (bases 1 to 2126)

AUTHORS Genoscope.

TITLE Direct Submission

JOURNAL Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail: seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)

COMMENT 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies, a division of Invitrogen.

FEATURES

Location/Qualifiers

source 1..2126

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="CS0DF037Y107"

/issue_type="Fetal brain"

/plasmid="pCMVSPORT_6"

ORIGIN

Alignment Scores:

Pred. No.: 3,346-108 Length: 2126

Score: 1907.50 Matches: 337

Percent Similarity: 99.70% Conservative: 0

Best Local Similarity: 99.70% Mismatches: 0

Query Match: 99.25% Indels: 1

DB: 3 Gaps: 1

US-10-791-017A-2_COPY_319_656 (1-338) x CR610888 (1-2126)

Qy 1 G1YGLYARGLYGLYVMECTGYSERALAGLYLUALARGLYGLYPHEANLYSPROGLYGLY 20

Db 943 GAGAGACGCGGTGATGGGC---GCTGAGAGACGAGTGGCTTCAATAGCCCTGTGCA 999

Qy 21 PrometAaPGLuGlyProAaApleuAaArgGlyProProValAaPProAaPGLuAaPser 40

Db 1000 CCCATGATGAGAGACCAATCTTATGAGGCCCACTGTATGATCAAGATGAAGACTCT 1059

Qy 41 AapAaSerAla1leYrYalGInGlyLeuAaAapSerValThrLeuAaApleuAla 60

Db 1060 GACAAACAGTCAATTTATGTACAAAGATTAATGACAGTGTGACTGATGATGATCTGGCA 1119

Qy 61 AapPhePheLeuGInGlyValValValValValValValValValValValValValVal 80

Db 1120 GACTTCTTTAAGCAGTGTGGGTTGTTAAGATGAACAGAAAGAACTGGGCAACCATGATC 1179

Qy 81 Hs1e1eYrLeuAaApleuGInGlyValValValValValValValValValValValVal 100

Db 1180 CACATCTACCTGACAAAGAAACAGAAAGCCCAAGGCAATGCCACAGTCTCTATGAA 1239

Qy 101 AapProProThra1aYalValAlaValAlaValAlaValAlaValAlaValAlaValAla 120

Db 1240 GACCCACCCCACTGCTCCAGGCTGCGTGGAAATGGTTGATGGAAAGATTTTCAAGGAGC 1299

Qy 121 LysLeuLysValSerLeuAlaArgLysLysProProMetLysSerMetLysArgLysLys 140

Db 1300 AAACCTTAAGTCTCCCTGCTCGAAGAAAGCTTCAATGAACAGTATGCGGGTGGCTG 1359

Qy 141 ProProArgGluGlyValArgGlyMetProProProLeuAaArgGlyValProGlyVal 160

Db 1360 CCAACCCCGTGAAGGAGGAGGCAATGCCACCACTCCGTGAGAGTCCAGAGGCGCAAGA 1419

Qy 161 GlyProGlyValProMetGlyValArgMetGlyValArgGlyValArgGlyValArgGly 180

```

Db      1420 GGTCTGGGGGACCCATGGGTGCATGGGAGCCGCGAGAGATGAGAGAGCTTCCTCCT 1479
QY      181  ProArGlyProArGlySerArGlyAanProSerGlyGlyGlyAanValGlnHisArg 200
Db      1480 CCAAGAGAGACCCCGGGGTTCCCGAGGAAACCTCTCGAGAGAGAAACGTCAGACCA 1539
QY      201  AlaGlyAaPTProGlnCySPProAanProGlyCyGlyAaAnGlnAanPheAaTTPArGThr 220
Db      1540 GCTGAGAGATCGGACAGTGTCCCATCCGGGTTGTGAAACCAAGACTTCGCTGAGAAACA 1599
QY      222  GlnCyAaAnGlnCyAaValAaAProLyAProGlnGlyPheLeuProProPhePro 240
Db      1600 GAGTCGAACCAAGTAAAGGCCCAAGAGCTTCTCCGACACCTTTCCGACC 1659
QY      241  ProGlyGlyAaPaRgGlyArGlyGlyGlyProGlyGlyGlyAaRgGlyGlyGlyLeu 260
Db      1660 CCGGGTGTGTATGTGTGAGAGAGTGGCCCTGTGTGACATGGGGAGAGAGAGGTGGCTTC 1719
QY      261  MetAaPaRgGlyGlyProGlyGlyMetPheArGlyGlyGlyAaRgGlyGlyGly 280
Db      1720 ATGATCGTGTGTGTGTCCCGTGAATGTTCAAGGTGTGCTGTGTGAGACAGAGGTGGC 1779
QY      281  PheArGlyGlyAaRgGlyMetAaPaRgGlyGlyPheGlyGlyGlyAaRgArGlyGlyPro 300
Db      1780 TTCCTGTGTGTGCGGGGACATGACCGAGGTGTGCTTGTGTGAGAGAGAGAGGTGGCTTC 1839
QY      301  GlyGlyProProGlyProLeuMetGlnGlnMetGlyGlyAaRgArGlyGlyAaRgGlyGly 320
Db      1840 GGGGGGCCCCCTGACCTTTGATGAGAACAGATGGAGAGAGAGAGAGAGAGAGAGAGAGAG 1899
QY      321  ProGlyAaMetAaPaRgGlyGlyGlnHisAaRgGlnGlyAaRgArGlyGlyPro 338
Db      1900 CTTGGAATAATGATTAAGGCGAGACCTCTCAGAGAGCGAGAGATCGGCTTAC 1953

```

```

RESULT 9
CR625247 2159 bp mRNA linear HTC 21-JUL-2004
LOCUS
DEFINITION
full-length cDNA clone CS0DE05F11 of Placenta of Homo sapiens
(human).
ACCESSION
CR625247 GI:50506054
VERSION
CR625247.1
KEYWORDS
HTC; CNS/LT cDNA.
SOURCE
Homo sapiens (human)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1. (bases 1 to 2159)
AUTHORS
Li, W.-B., Gruber, C., Jessee, J. and Polayes, D.
TITLE
Full-length cDNA libraries and normalization
JOURNAL
Unpublished
REMARK
Contact: Feng Liang Email: fliang@lifetech.com URL:
http://fulllength.invitrogen.com/Invitrogen Corporation 1600
Faraday Avenue
2 (bases 1 to 2159)
REFERENCE
Genoscope.
AUTHORS
Direct Submission
TITLE
Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage:
JOURNAL
BP 131 91006 EVRY cedex - FRANCE (E-mail: seqref@genoscope.cns.fr
- Web: www.genoscope.cns.fr)
COMMENT
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by Life Technologies, a
division of Invitrogen.
FEATURES
Location/Qualifiers
1. 2159
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon.9606"
/clone="CS0DE05F11"
/libase_type="Placenta"
/plasmid="pCMVSPORT_6"

```

```

FEATURES
Location/Qualifiers
1. 2159
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon.9606"
/clone="CS0DE05F11"
/libase_type="Placenta"
/plasmid="pCMVSPORT_6"

```

ORIGIN

Alignment Scores:

```

Pred. No.: 3 4e-108 Length: 2159
Score: 1907.50 Matches: 337
Percent Similarity: 99.70% Conservative: 0
Best Local Similarity: 99.70% Mismatches: 1
Query Match: 99.25% Indels: 1
DB: 3 Gaps: 1

```

US-10-791-017a-2_COPY_319_656 (1-338) x CR625247 (1-2159)

```

QY      1  GilyAaRgGlyGlyMetGlySerAaGlyGlyAaRgGlyGlyPheAaAnLySProGlyGly 20
Db      1000 GGAGAGCGGGGTGATGGGC---CGTGAAGAGCGAGGTGGCTTCAATTAAGCCTGGTGA 1056
QY      21  PrometAaRgGlnGlyProAaPheAaPheGlyProProValAaSProAaRgGlyAaPaR 40
Db      1057 CCCATGATGAAGAGACCAATCTTGTATCTAAGGCCCACTGTAGATCCAGATGAAGACTCT 1116
QY      41  AaPaRSerAaIleTyValGlnGlyLeuAaAaSPSerValThrLeuAaPheAaAa 60
Db      1117 GACACAGTGCATTTATATATCAAGAGATTAAATGACAGTGTACTTCAATGATCTGGCA 1176
QY      61  AaPhePheGlyGlnCyGlyValVallyAaMetAaAnLyAaRgThrGlyGlnProMetIle 80
Db      1177 GACTCTTTAAGACAGTGTGGGTGTGTTAGATGAACAAGAGAGAGAGAGAGAGAGAGAGAG 1236
QY      81  HisIleTyLeuAaPlyGlnThrGlyGlyGlyProLyAaPheAaAaThrValSerTyrgly 100
Db      1237 CACATCTTACCTGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1296
QY      101  AaSPProThrAlaValAaAaValGlnTTPheAaPheGlyGlyAaPheGlnGlySer 120
Db      1297 GACCAACCACTGCCAGAGCTGCCGTGGAATGTTGATGGAGAAATTTCAAGGAGAC 1356
QY      121  LyLeuLyValSerLeuAaRgLyAaRgLyAaPheAaPheAaPheAaPheAaPheAaPheAa 140
Db      1357 AAACCTTAAGTCTCCCTTCTCGAAGAGAGCTCCAAATGAACAGTATCGGGGTGTCTG 1416
QY      141  ProProAaRgGlnGlyAaRgGlyMetProProProLeuAaRgGlyGlyProGly 160
Db      1417 CCAACCCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1476
QY      161  GlyProGlyGlyProMetGlyAaRgMetGlyGlyAaRgGlyGlyAaPaRgGlyGlyPhePro 180
Db      1477 GGTCTGGGGGACCCATGGGTGCATGGAGGCCGCGAGAGAGATGAGAGAGCTTCCTCCT 1536
QY      181  ProArGlyProArGlySerArGlyAanProSerGlyGlyGlyAanValGlnHisArg 200
Db      1537 CCAAGAGAGACCCCGGGGTTCCCGAGGGAACCTCTGAGAGAGAGAGAGAGAGAGAGAGAG 1596
QY      201  AlaGlyAaPTProGlnCySPProAanProGlyCyGlyAaAnGlnAanPheAaTTPArGThr 220
Db      1597 GCTGAGAGATCGGACAGTGTCCCAATCCGGTGTGTGAAACCAAGACTTCGCTGAGAAACA 1656
QY      221  GlnCyAaAnGlnCyAaValAaAProLyAProGlnGlyPheLeuProProPhePro 240
Db      1657 GAGTCGAACCAAGTAAAGGCCCAAGAGCTTCTCCGACACCTTTCCGCC 1716
QY      241  ProGlyGlyAaPaRgGlyArGlyGlyGlyProGlyGlyGlyAaRgGlyGlyGlyLeu 260
Db      1717 CCGGGTGTGTATGTGTGAGAGAGTGGCCCTGTGTGACATGGGGAGAGAGAGAGAGAGAG 1776
QY      261  MetAaPaRgGlyGlyProGlyGlyMetPheArGlyGlyGlyAaRgGlyGlyGly 280
Db      1777 ATGATCGTGTGTGTGTCCCGTGAATGTTCAAGGTGTGCGGTGTGAGACAGAGGTGGC 1836
QY      281  PheArGlyGlyAaRgGlyMetAaPaRgGlyGlyPheGlyGlyGlyAaRgArGlyGlyPro 300
Db      1837 TTCCTGTGTGTGCGGGGACATGACAGAGGTGTGCTTGTGTGAGAGAGAGAGAGAGAG 1896
QY      301  GlyGlyProProGlyProLeuMetGlnGlnMetGlyGlyAaRgArGlyGlyAaRgGlyGly 320

```

DB	1997	GGGGGGCCCCCTTGACCTTTATATGGAACAGATGGAGAGAAAGAGAGACCTGGAGGA	1956
By	321	ProGlyLyMetCAspLYGlyGluHieArgGlnGluArgaGAspArgProTyr	338
Db	1957	CCTGGAAAAATGATATTAAGGGAGACACCGTACAGAGCCGACAGATCGGCCCTTAC	2010
RESULT 10			
LOCUS	AK019460	2107 bp	linear
DEFINITION	Mus musculus 18 days pregnant adult female placenta and extra embryonic tissue cDNA, RIKEN full-length enriched library, clone:3830417B11 product:Erwing sarcoma homolog, full insert sequence.		
ACCESSION	AK019460		
VERSION	AK019460.1	GI:12859676	
KEYWORDS	HTC, CAP trapper.		
SOURCE	Mus musculus (house mouse)		
ORGANISM	Mus musculus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
AUTHORS	Carninci, P. and Hayashizaki, Y.		
TITLE	High-efficiency full-length cDNA cloning		
JOURNAL	Meth. Enzymol. 303, 19-44 (1999)		
MEDLINE	99279253		
PUBMED	10349656		
REFERENCE	2		
AUTHORS	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.		
TITLE	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes		
JOURNAL	Genome Res. 10 (10), 1617-1630 (2000)		
MEDLINE	20499374		
PUBMED	11042159		
REFERENCE	3		
AUTHORS	Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Kono, H., Akiyama, J., Nishi, K., Kitenaka, T., Tashiro, H., Itoh, M., Suni, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaue, S., Ikegami, T., Kashiwagi, K., Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Matsumi, Y., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.		
JOURNAL	RIKEN integrated sequence analysis (RISA) system-384-format		
MEDLINE	Genome Res. 10 (11), 1757-1771 (2000)		
PUBMED	20530913		
REFERENCE	4		
AUTHORS	The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.		
TITLE	Functional annotation of a full-length mouse cDNA collection		
JOURNAL	Nature 409, 685-690 (2001)		
MEDLINE	5		
PUBMED	The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.		
REFERENCE	Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs		
AUTHORS	Nature 420, 563-573 (2002)		
TITLE	6 (bases 1 to 2107)		
JOURNAL	Adachi, J., Aizawa, K., Akahira, S., Akimura, T., Arai, A., Aono, H., Aizawa, T., Bono, H., Carninci, P., Fukuda, S., Fukunishi, Y., Furuno, M., Hanagaki, T., Harada, A., Hayatsu, N., Hiramoto, K., Hiraoaka, T., Hori, P., Imotani, K., Ishii, Y., Itoh, M., Izawa, M., Kanakawa, T., Kato, H., Kawai, J., Kojima, Y., Kono, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Okazaki, Y., Okido, T., Owa, C., Saito, H., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D., Shibata, K., Shibata, Y., Shingawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Tanaka, T., Tejima, Y., Toyai, T., Yamamura, T., Yasunishi, A., Yoshida, K., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.		
REFERENCE	Direct Submission		

[illegible]

GAGAGAGATTCTCGAGTTAATTAATCATCCCCCCCCCCC 3'. cDNA was cleaved
with XhoI and SctI. Cloning sites, 5' end: XhoI; 3' end: SctI.
Host: SOLR.

FEATURES

source

Location/Qualifiers

1..1379
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM DB:3300002D11"
/db_xref="taxon:10090"
/clone="3300002D11"
/issue_type="head"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="17 days embryo"
1..1004
/note="unnamed protein product; Ewing sarcoma homolog
(MGI:99960, GB|NM_007968, evidence: BLASTN, 99%,
match=2172)
putative"

CDS

/codon_start=3
/protein_id="BAB29301.1"
/db_xref="GI:12852168"
/translation="LGSAGERGGRKRGPHMDEPDLDLGLPIDPEDSDNSAIYVQ
LNDNTLDDLDLADFFKOCGVYKMKRGTQPMIHYLDKRTGPKDVTASVEDPTAKA
AVEMFDGKDFOGSKLKVSLARKKPPMNSMRGMPPRRGGMPPRLGARGGPGGPGP
MGMGRGCDRGSPPRGPRGSKGNPSSGAGVGRAGDMQCPNPGGAGNONTARTCN
QCAVPRGEPFLPPPPGDRGGGGRGGGLMDGGGGMFRGGGDRGGF
RGGGMDRGFGGGRGGPGPPPLMEQGRGRRGGGGRMDKSHHRRDRPY"
1351..1356
/note="putative"
1379
/note="putative"

ORIGIN

Alignment Scores:
Pred. No.: 2,47e-106 Length: 1379
Score: 1874.00 Matches: 328
Percent Similarity: 99.70% Conservative: 4
Best Local Similarity: 98.50% Identical: 1
Query Match: 97.50% Gaps: 0
DB: 3

US-10-791-017a-2_COPY_319_656 (1-338) x AK034366 (1-1379)

QY 6 MetGlySerAlaGlyLuarGlyGlyPheAsnLysProGlyGlyProMetAspGlyGly 25
DB 3 CTGGGCGAGCGCTCGAGAGCGAGGCTTCAATTAAGCTGTGAGACCATGATGAAGA 62
QY 26 ProAspLeuAspLeuGlyProProValAspProAspGlyAspSerAspAsnSerAla 45
DB 63 CCAGATCTGATTAAGCTCTTCTATATGATCCGATGAACCTGACCAACATGCAATT 122
QY 46 TyrValGlnGlyLeuAsnAspSerValThrLeuAspLeuAlaAspPhePheGln 65
DB 123 TATGTCCAAGGATTAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 182
QY 66 CysGlyValValIysMetCAsnLysValGlnProMetIleHisIleTyrLeuAsp 85
DB 183 TGTGGGGTGTCTCAAGATGAACAAGAACTGACCAACCATGATCATATCTACTGGAT 242
QY 86 LysGluThrGlyLysProLysGlyAspAlaThrValSerTyrGluAspProProThrAla 105
DB 243 AAGGACACAGAAAGCTTAAGGGGAGCGCACGTCCTCATTAAGATCCACCACTGCA 302
QY 106 LysAlaAlaValGluTyrPheAspGlyLysAspPheGlnGlySerIleLeuLysValSer 125
DB 303 AAGGCTCGCGTGAAGGTTTGAATGGAAGATTTTCAAGAGCACTTAAGCTCTCT 362
QY 126 LeuAlaArgLysLysProProMetAsnSerMetArgGlyGlyLeuProProArgGlyGly 145
DB 363 CTGTCCCGAAGAAAGCTTCAAGAAACAGCATCGGGAGGAGCATGCACCTCGTGAAGG 422
QY 146 ArgGlyMetProProLeuArgGlyGlyProGlyGlyProGlyGlyProGlyGlyPro 165

DB 423 AGGGGATGCCACACACCTTCTGAGGTCTGTGTGCCCCAGAGGCTTGAAGACCC 482
QY 166 MetGlyArgMetGlyGlyArgGlyGlyAspArgGlyGlyPheProProArgGlyProArg 185
DB 483 ATGGGTGCGATGGGTGGCGGTGAGAGAGACAGAGGGGGCTTCTCTCCAAAGAGGCCCGA 542
QY 186 GlySerArgGlyLysProSerGlyGlyGlyLysValGlnHisArgAlaGlyAspTyrGln 205
DB 543 GGTCCAGAGAAACCTCTTGAGAGAGAAATGTCAGAGACCGAGCTGAGACTGGCAG 602
QY 206 CysProAspProGlyCysGlyLysAsnGlnAsnPheAlaTyrArgThrGluCysAsnGlnCys 225
DB 603 TGTCCATCCGGGCTGTGAAACAGAACTGCTTGAAGAACAGATCAACCAATGT 662
QY 226 LysAlaProLysProGlyGlyPheLeuProProProPheProProGlyGlyAspArg 245
DB 663 AAGGCCCTTAAGCCGAGGGCTTCTCTCCGACCCCTTTCACCTCCGGGTGATCGT 722
QY 246 GlyArgGlyGlyProGlyGlyMetArgGlyGlyLysArgGlyGlyLeuMetAspArgGly 265
DB 723 GAGCGAGGTGGCCCTGTGTCATGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 782
QY 266 ProGlyGlyMetPheArgGlyGlyArgGlyGlyLysAspArgGlyGlyPheArgGlyArg 285
DB 783 CTGAGAGAAATTCAGAGGTGCGAGAGGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 842
QY 286 GlyMetAspArgGlyGlyPheGlyGlyGlyArgArgGlyGlyProGlyGlyProGly 305
DB 843 GGAATGAGCCGAGGTGGCTTGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 902
QY 306 ProLeuMetGlyGlnMetGlyGlyArgArgGlyGlyArgGlyGlyProGlyLysMetAsp 325
DB 903 CTTTAAATGACACGATGAG 962
QY 326 LysGlyGlyHisArgGlnGluArgArgAspArgProTyr 338
DB 963 AAGGCGAGACCGTCAAG 1001
RESULT 12
AK034755
LOCUS
DEFINITION
AK034755 2373 bp mRNA linear HTC 03-Apr-2004
MUS MUSCULUS 12 days embryo embryonic body between diaphragm region
and neck cDNA, RIKEN full-length enriched library, clone:943003IM11
product:Ewing sarcoma homolog, full insert sequence.
ACCESSION
AK034755
VERSION
AK034755.1 GI:26084174
SOURCE
HTC; CAP trapper.
MUS MUSCULUS (house mouse)
ORGANISM
MUS MUSCULUS
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1
REFERENCES
AUTHORS
Carninci, P. and Hayashizaki, Y.
TITLE
High-efficiency full-length cDNA cloning
JOURNAL
METH. ENZYMO. 303, 19-44 (1999)
MEDLINE
99279253
PUBMED
10349636
2
REFERENCES
AUTHORS
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
Itoh, M., Komano, H., Okazaki, Y., Mutamatsu, M. and Hayashizaki, Y.
TITLE
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL
GENOME RES. 10 (10), 1617-1630 (2000)
MEDLINE
20499374
PUBMED
11042159
3
REFERENCES
AUTHORS
Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Saeki, N., Carninci, P.,
Komano, H., Akiyama, J., Nishi, K., Kitsuai, T., Taahiro, H., Itoh, M.,
Sumi, N., Iehi, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A.,
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Matsubuchi, M.,
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,

TITLE Okazaki, Y., Muramatsu, M., Inoue, Y., Kita, A. and Hayashizaki, Y.
 RIKEN Integrated sequence analysis (RISA) system-384-format
 JOURNAL sequencing pipeline with 384 multicapillary sequencer
 MEDLINE Genome Res. 10 (11), 1757-1771 (2000)
 20530913
 PUBMED 11076861
 REFERENCE 4
 AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and the
 PANTOM Consortium.
 TITLE Functional annotation of a full-length mouse cDNA collection
 JOURNAL Nature 409, 685-690 (2001)
 REFERENCE 5
 AUTHORS The PANTOM Consortium and the RIKEN Genome Exploration Research
 Group Phase I & II Team.
 TITLE Analysis of the mouse transcriptome based on functional annotation
 JOURNAL of 60,770 full-length cDNAs
 REFERENCE 6 (bases 1 to 2373)
 AUTHORS Nature 420, 563-573 (2002)
 REFERENCE 6
 AUTHORS Adachi, J., Aizawa, K., Akimura, T., Arikawa, T., Bono, H., Carninci, P.,
 Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W.,
 Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T.,
 Horii, P., Imotani, K., Ishii, Y., Itoh, M., Kagawa, T., Kasukawa, T.,
 Katoh, H., Kawai, J., Kohji, Y., Kondo, S., Kono, H., Kouda, M.,
 Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,
 Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N.,
 Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,
 Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,
 Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,
 Takekida, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,
 Muramatsu, M. and Hayashizaki, Y.
 TITLE Direct Submission
 JOURNAL Submitted (16-JUN-2001) Yoshihide Hayashizaki, The Institute of
 Physical and Chemical Research (RIKEN), Laboratory for Genome
 Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
 RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
 Kanagawa 230-0045, Japan (E-mail: genome-res@gscc.riken.jp,
 URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222,
 Fax: 81-45-503-9216)
 COMMENT cDNA library was prepared and sequenced in Mouse Genome
 Encyclopedia Project of Genome Exploration Research Group in Riken
 Genomic Sciences Center and Genome Science Laboratory in Riken.
 Division of Experimental Animal Research in Riken contributed to
 prepare mouse tissues.
 Please visit our web site for further details.
 URL: http://genome.gsc.riken.jp/
 URL: http://fantom.gsc.riken.jp/
 FEATURES
 source Location/Qualifiers
 1. .2373
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="FANTOM:DB:943003IM11"
 /db_xref="taxon:10090"
 /clone="943003IM11"
 /tissue_type="embryonic body between diaphragm region and
 neck"
 /clone_1lb="RIKEN full-length enriched mouse cDNA library"
 /dev_stage="12 days embryo"
 43. .2011
 /note="Ewing sarcoma homolog (MCD|MGI:99960, GB|NM_007968,
 evidence: BLASTN, 99%, match=2172)
 putative"
 2358. .2363
 /note="putative"
 2373
 /note="putative"
 ORIGIN
 polyA_site
 polyA_site
 /note="putative"
 Alignment Scores:
 Pred. No.: 6.36e-106 Length: 2373
 Score: 1871.50 Matches: 330
 Percent Similarity: 98.82% Conservative: 4
 Best Local Similarity: 97.63% Mismatches: 3

Query Match: 97.37% Indels: 1
 DB: 3 Gaps: 1
 US-10-791-017a-2_copy_319_656 (1-338) x AK034755 (1-2373)
 QY 1 G1G1YATGAG1G1YMeC1YSeR1aG1YUATG1YG1YPhEaNs1YSeR1YProG1Y 20
 DB 998 GGAAGACCGCGGTGACCTGGCC---GCTGGAGAGCGAGGTGGCTTCAATAGCCTGTGGA 1054
 QY 21 ProMetAspGluGlyProAspLeuAspLeuGlyProProValAspProAspGluAspSer 40
 DB 1055 CCCATGATGAGAGACCAAGATCTTATGATCAAGCCCTTCATGATGCCATGAAAGCTCT 1114
 QY 41 AspaNsErAla1eTyVa1G1nG1YLeuAaNsPSeVa1ThrLeuAspLeuAla 60
 DB 1115 GACACAGTGCATTTATGTCAGAGATTTAATGACATGATCACTGATGATCTGGCA 1174
 QY 61 AspPhePheGlnCysG1YVal1YallysMeCaNs1YsArgThrG1Yn1nProMetIle 80
 DB 1175 GACTCTTTAAGCAGGTGGGTGTTCAGATGAAACAAGAACTGACCAACCATGATC 1234
 QY 81 His1eTyTrLeuAsp1YsG1UthrG1YlybPro1YsG1YAspAlaThrValSerTyGlu 100
 DB 1235 CATATCTACTGATTAAGAGAGACAGAAAGCCTTAAAGGGAGACCAAGTCTCTATGAA 1294
 QY 101 AspProProThrAla1YsAla1aVala1G1UTpPheAspG1YlybAspPheGln1Yser 120
 DB 1295 GATCCACCACTGACCAAGGCTCCGTGATGATGTTGATGGAGAAATTTCAAGAAAGC 1354
 QY 121 YsLeu1YsVal1Ser1eUa1aArg1Ys1YsProProMetAsn1Se1MetArg1Yg1YLeu 140
 DB 1355 AAACCTTAAAGTGTCTTTCGCCGAAAGAGCCTCCATGAACGATGCGGGAGGACATG 1414
 QY 141 ProProArgGluG1YAspG1YmetProProProleuAspG1YG1YProG1YProG1Y 160
 DB 1415 CCACTCCGAGAGGAGGAGGAGATGACCAACCACTTCGAGAGTCTGTGTGCGCCAGGA 1474
 QY 161 G1YProG1YG1YProMeC1YArgMeC1YG1YAspArg1YG1YAspArg1YG1YPhePro 180
 DB 1475 GGCCTGAGAGACCCATGAGTGTGATGAGAGCCCTGAGAGAGACAGAGGGCTTCCT 1534
 QY 181 ProArgG1YProArgG1YSeRArg1YAsnProSeRArg1YG1YAsnValGln1aArg 200
 DB 1535 CCAAGAGGAGCCCGAGGCTCCAGAGGAAACCCCTCGAGAGAGAAATGCCAGCACCGA 1594
 QY 201 AlaG1YAspTrG1nCybAspAsnProG1YCYG1YAsnG1YAsnPheAlaTrpAArgTr 220
 DB 1595 GCTGGAGACTGGCAGTGTCCCAATCAGGGCTGTGAAACCAAGAACTTCCTGGAGAAACA 1654
 QY 221 G1UCyAsnG1nCyb1YsAla1aPro1YsProG1nG1YPhEaNsProProPheProPro 240
 DB 1655 GAATGCAACCAATGTAAAGCCCTTAAGCCCGAGGGCTTCCTCCGCGACCTTCCACCT 1714
 QY 241 ProG1YG1YAspArg1YAspArg1YProG1YG1YMeCArgG1YG1YArgG1YG1YLeu 260
 DB 1715 CCGGAGTGTGATGTGAGAGAGGTGGCCCTGTGGATCGAGAGAGAGAGAGAGAGAGAGAG 1774
 QY 261 MetAspArgG1YG1YProG1YG1YMetPheArgG1YG1YArgG1YG1YAspArgG1YG1Y 280
 DB 1775 ATGACACCGTGTGTCTCTGAGAGAAATGTTCAAGGTGGAGAGAGTGAAGACAGAGAGAGC 1834
 QY 281 PheArgG1YG1YArgG1YMeCAspArgG1YG1YPheG1YG1YArgArgG1YG1YPro 300
 DB 1835 TTCCGAGGTGGCCGTGATGTATCCAGATGCTTCCTGTGGAGAGAGAGAGAGAGAGAGAG 1894
 QY 301 G1YG1YProProG1YProleuMetG1nMeC1YG1YArgArgG1YG1YArgG1YG1Y 320
 DB 1895 GGGGGGCTCTCGACCTTTAATGAAACAGATGGAGAGAAAGAGAGAGAGAGAGAGAGAGAG 1954
 QY 321 ProG1YlyMetAsp1YsG1YUth1aArgG1nG1YArgArgAspArgProTyR 338
 DB 1955 CCTGGGAAATGATTAAGCGAGGACCGCTCAGAAACGAGAGACCGGCGCTTAC 2008

RESULT 13
CR604639
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
REMARK
REFERENCE
AUTHORS
TITLE
JOURNAL
REMARK
COMMENT
FEATURES
source
ORIGIN
Alignment Scores:
Pred. No.:
Score:
Percent Similarity:
Best Local Similarity:
Query Match:
DB:
US-10-791-017a-2_COBY_319_656 (1-338) x CR604639 (1-2236)

CR604639 2236 bp mRNA linear HTC 21-JUL-2004
full-length cDNA clone CS0DU013YN08 of T cells (Jurkat cell line)
Cot 10-normalized of Homo sapiens (human).
CR604639
CR604639.1 GI:50485446
HTC; cNSLT cDNA.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 2236)
L.I.W.B., Gruber, C., Jesse, J. and Polyes, D.
Full-length cDNA libraries and normalization
unpublished
Contact : Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Faraday Avenue
2 (bases 1 to 2236)
Genoscope.
Direct Submision
Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by Life Technologies, a
division of Invitrogen.
Location/Qualifiers
1..2236
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DU013YN08"
/cissue_type="T cells (Jurkat cell line) Cot
10-normalized"
/plasmid="pCMVSPORT_6"

4,32e-98 Length: 2236
1744.50 Matches: 314
92.90% Conservation: 0
92.90% Mismatches: 1
90.76% Indels: 24
Gaps: 2

1 G1G1YARGLYGLYMETGLYSERIALAG1GLYUARG1GLYPHEASNTYSPROGLYGLY 20
973 GGAAGACGGCGGTGGAATGGGC---GCTGAGACGCGAGTGGCTTCAATTAAGCTGTGGA 1029
21 PROMETASPG1GLYPROASPLEUASPLEUGLYPROPROVALASPPROASPG1UASPSER 40
1030 CCCATGATGAGAGACCAAGATCTTGTAGAGCCCACTGTAGATCCAGATGAAGACTCT 1089
41 AAPAASERIALIETRYVALG1NG1YLSAASASPSERVALTHIRLEUASPSPLEUALA 60
1090 GACCAACAGTGCATTTATATACAGAGATTAAAGACAGTGTACTAGATGATCTGGCA 1149
61 AAPPHepheleG1ncYsG1yValVal1ySMetAsn1yARgThrg1yG1nPRomEctile 80
1150 GACTTCTTTAAAGCAAGTGGGTGTGTTAAGATGACAAAGAACTGGGCAACCATGATC 1209
81 Hs11eTyTLeuAspiYsG1uThrg1yLysPRoLySg1yAspAlaThrValSerTyrg1u 100
1210 CACATCTACTGTGACAAAGAAACAGAGAAACCCAAAGGCGATGCCCACTGTCCTATGAA 1269
101 AAPPRoTThrAla1yVal1a1aVal1G1uTTPPhEASpG1yLeaSPheG1nG1ySer 120

Db 1270 GACCCACCACTGCCAAGCTGCCGTGGAATGCTTTGATGGGAAATTTTCAAGGAGC 1329
QY 121 LysLeu1yValSerLeu1a1a1yLysPProMetAsnSerMetA1rg1yG1yLeu 140
Db 1330 AAACCTTAAGTCTCCCTTCTCGAAGAAAGCTCCATATGAACGATATGCGGGGTGCTG 1389
QY 141 PROPRARGL1GLYVRGLYMETPROPRORPROLEUARG1GLY1YPROGLY1YPROGLY 160
Db 1390 CCACCCGTGAGG3CAGAGGCATATGCCACCACTCGTGGAGAGTCCAGAGGCCACAGA 1449
QY 161 GLYPRG1YGLYPROMETGLYARMETGLY1YARG1YGLYASPARGLY1YPhEPro 180
Db 1450 GGTCTGGGGGACCATGGTGGCATGGAGGCCGTGGAGAGATGAGAGGCTTCCCT 1509
QY 181 PROARG1YPROARG1YSEAR1YASMPROSEAR1YGLY1YANVALG1NH1BAR 200
Db 1510 CCAGAGAGACCCCGGGGTTCCGAGGAAACCCCTTGAGAGAGAAACGTCAGACCCA 1569
QY 201 A1AG1YASPTTGL1ncYsPAsnPRoAsnPRoGLYCYsG1yAsnG1nAsnPhEalATTPATh 220
Db 1570 GCTGAGACTGCGCAGTGTCCCAATCCGGGTGTGGAAACAGAACTTCGCTTGAGACAA 1629
QY 221 GLUCYASnG1ncYsLys1a1aPProLYsPProGL1YPhE1uPProPProPProPPro 240
Db 1630 GAGTGCACCAAGTGTAAAGCCCAAGCTTGAAGGCTTCTCCGCAACCTTTCCGCC 1689
QY 241 PROGLY1YASPARGLY1YARG1YGLY1YPROGLY1YMETARG1YGLY1YARG1YGLY 260
Db 1690 CCGGCGTGTGATCGTGTGAGAGGTGGCCCTGGTGCATCGGGGAGAAAGAGGTGGCCT 1749
QY 261 METASPARGLYGLYPROGLY1YMETPHEARG1Y1YARG1YGLY1YASPARGLY1Y 280
Db 1750 ATGGAATCGTGTGTGTCCCGGTGAATGTTCAGAGGTGGCCGTGTGAGACAGAGGTGGC 1809
QY 281 PHEARG1YGLY1YARG1YMETASPARGLY1YPhEGLY1YGLY1YARG1YGLY1YPRO 300
Db 1810 TTCGGTGTGGCCGGGCAATGACCAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1869
QY 301 G1YGLY1YPROGLY1YPROLEUWETGLUG1nMetG1y1YARG1YGLY1YARG1YGLY 320
Db 1870 GGGGGGCC-CCA----- 1880
QY 321 PROGLY1YMETASPLYSGLY1YHISARG1GLYUARG1YASPARGLY1YPROTYR 338
Db 1881 -----GGCGAGCACCGTCAAGAGCGCAGAGATCGCCCTTAC 1916

RESULT 14
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED

AK049743 2269 bp mRNA linear HTC 03-APR-2004
Mus musculus 12 days embryo spinal cord cDNA, RIKEN full-length
enriched library, clone: C530046A18 product: bwing sarcoma homolog,
full insert sequence.
AK049743
AK049743.1 GI:26093614
HTC; CAP trapper.
Mus musculus (house mouse)
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sclirognathi; Muridae; Murinae; Mus.
1 Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning
Meth. Enzymol. 303, 19-44 (1999)
99279253
10349636
2 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)
20499374
11042159

Accession	Protein	Position	Sequence	Position
Oy	242	-----G	YGLVLAAPARGGLVARGGLVGLIYProGLVGLIYmeAARGGLVGLIYArg	257
Db	4223	ATTCTCACCTT	AGGTGGTGTATCGTGCGCAGAGGTGGCCCTGTGTGGATATCGGGAGGAGGA	4282
Oy	258	G	YGLIYLeuMeAaPARGGLVGLIYProGLIYmeC PheaArgGLVGLIYARGGLVGLIYAaP	277
Db	4283	GGTGGCCCTCA	TGATCGATCGTGTGTGTCCTCCGTTGGAAATTTCAAGAGTGCCCGTGTGTGAGAC	4342
Oy	278	A	RGGLYGLIYpheaArgGLVGLIYARGGLYmeAaPARGGLVGLIYpheaGLVGLIYARGArg	297
Db	4343	AGAGGTGCTTCC	GGTGGTGGCCGGGCGATGAGCCGAGGTGGCTTTGGTGTGAGGAAGAGA	4402
Oy	298	G	LYGLIYProGLIYGLIYProProGLIYProLeuMeCGLIYInMeCGLIYGLIYARGArgGLVGLIY	317
Db	4403	GGTGGCCCTG	TGGGGGGCCCCCTTGACCTTTGATGTGAACAGATGGGAGGAAGAGGAGGA	4462
Oy	318	A	RGGLYGLIYProGLIYLeuMeAaPArgGLVGLIYInSAARGGLInGLIYARGArgAaPARGPro	337
Db	4463	CGTGGAGGACCT	TGGAAAAATGATTAAGCCAGACACCTGTACAGAGCGCAGAGATCGGCC	4522
Oy	338	Tyr	338	
Db	4523	TAC	4525	

Search completed: February 21, 2005, 04:12:05
Job time : 2560 secs

APPLICANT: OTSUKA, MOTOKUKI
 APPLICANT: NAGAHART, KENJI
 APPLICANT: MASUHO, YASUHIKO
 TITLE OF INVENTION: NOVEL FULL-LENGTH cDNA
 FILE REFERENCE: 084335/0160
 CURRENT APPLICATION NUMBER: US/10/094,749
 CURRENT FILING DATE: 2002-03-12
 PRIOR APPLICATION NUMBER: 60/350,435
 PRIOR FILING DATE: 2002-01-24
 PRIOR APPLICATION NUMBER: JP 2001-328381
 PRIOR FILING DATE: 2001-09-14
 NUMBER OF SEQ ID NOS: 3381
 SOFTWARE: PatentIn Ver. 2.1
 SEQ ID NO 1074
 LENGTH: 1988
 TYPE: DNA
 ORGANISM: Homo sapiens
 US-10-094-749-1074

Alignment Scores:
 Pred. No.: 4e-155 Length: 1988
 Score: 1922.00 Matches: 338
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 17 Gaps: 0

US-10-791-017a-2_COPY_319_656 (1-338) x US-10-094-749-1074 (1-1988)

QY 1 G1YGYAARG1YGLYMETG1YSER1AG1YGLUARG1YGLYPHEANLYSPROGLYGLY 20
 DB 798 GGAGGAGCGCGTGGAAATGGGCGAGCGTGGAGAGAGGTGGCTTCAATTAAGCGTGGGA 857
 QY 21 PROMETASPG1UGLYPROAEPLEUASPLEUGLYPROFOVALASPPROAEPGLUASPSER 40
 DB 858 CCCATGATGAAGAGCAGATCTTGATCTGATCTGATCTGATCTGATCTGATCTGATCTGAT 917
 QY 41 ASAPSASER1A1ETRYVALING1YLEUASAPSASER1A1ETRYVALING1YLEUASAPS 60
 DB 918 GACACAGTGCATTTAT 977
 QY 61 ASPPHEPHELYSG1YCG1YVAL1YAL1YMETASNL1YARG1YGLINPROMET1LE 80
 DB 978 GACTCTTTAAGCAGTGTGGGTGTTTAAATGAACAAGAACTGGGCAACCATGATC 1037
 QY 81 H1E1ETRYLEUASPLYVGLU1YTH1YLYSPROLYSG1YASPA1ATHTVALSER1YGLU 100
 DB 1038 CACATCTACCTGGAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 1097
 QY 101 ASPPROPTH1A1ALYVAL1A1A1VAL1GLUTPH1EASPG1YLYSASPHEGLIN1YSER 120
 DB 1098 GACCCACCTGCAAGCTGCAAGCTGCAAGCTGCAAGCTGCAAGCTGCAAGCTGCAAGCTG 1157
 QY 121 LYSLEU1YVAL1SER1EUAL1AARG1YLYSPROF1EASPSER1EASPSER1EASPSER 140
 DB 1158 AATTTAAAGTCTCTCTGCTGCAAGAGCTCCATTAAGCACTTAAGCGGGGCTGCTG 1217
 QY 141 PROB1EASPG1UGLYAARG1YMET1PROB1EASPG1UGLYAARG1YMET1PROB1EAS 160
 DB 1218 CCAACCCGTCAGGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1277
 QY 161 G1YVPROGLYGLYPROMETG1YVARG1YGLYVARG1YGLYVARG1YGLYVARG1YGLY 180
 DB 1278 GGTCTTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1337
 QY 181 PROARG1YPROARG1YSEAR1YASNP1EASPSER1YGLYGLYVARG1YGLYVARG1YGLY 200
 DB 1338 CCAAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1397
 QY 201 ALA1YASPT1P1G1YVARG1YGLYVARG1YGLYVARG1YGLYVARG1YGLYVARG1YGLY 220
 DB 1398 GCTGAGAGAGTGGAGTGTCCCAATCCGGGTTGTGAAACCAAGAACTTCCGCTGAGAGAA 1457

QY 221 GLUCYASANG1YCYEL1YVAL1EAP1EASPSER1YGLYVARG1YGLYVARG1YGLYVARG 240
 DB 1458 GACTGACACCAAGTGTAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1517
 QY 241 PROGLYV1YASPA1YARG1YGLYVARG1YGLYVARG1YGLYVARG1YGLYVARG1YGLY 260
 DB 1518 CCGGAGTGTGATCTGTCAGAGAGTGGCTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1577
 QY 261 METASPA1YGLYVARG1YGLYVARG1YGLYVARG1YGLYVARG1YGLYVARG1YGLYVARG 280
 DB 1578 ATGATCTGTGTGTCTCCGAGTGAATGTTCAAGAGTGGTGGTGGTGGTGGTGGTGGTGG 1637
 QY 281 PHEARG1YGLYVARG1YMETASPA1YGLYVARG1YGLYVARG1YGLYVARG1YGLYVARG 300
 DB 1638 TTCCGT 1697
 QY 301 G1YGLYVARG1YGLYVARG1YGLYVARG1YGLYVARG1YGLYVARG1YGLYVARG1YGLY 320
 DB 1698 GGGGGGCCCCCTGGAGCTTTGATGAGAACAGATGGAGAGAGAGAGAGAGAGAGAGAG 1757

RESULT 2

US-09-822-830A-49/c
 Sequence 49, Application US/09822830A
 Patent No. US20020142952A1

GENERAL INFORMATION:
 APPLICANT: Genetics Institute, Inc.
 APPLICANT: Wong, Gordon G.
 APPLICANT: Clark, Hilary
 APPLICANT: Rechele, Kim
 APPLICANT: Agostino, Michael J.
 APPLICANT: Howes, Steven H.
 APPLICANT: Resnick, Richard J.
 APPLICANT: Gulukota, Kamalakara
 APPLICANT: Graham, James R.
 TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING NOVEL SECRETED PROTEINS
 FILE REFERENCE: GIN 6402
 CURRENT APPLICATION NUMBER: US/09/822,830A
 PRIOR FILING DATE: 2001-03-29
 PRIOR APPLICATION NUMBER: 60/195,604
 PRIOR FILING DATE: 2000-04-06
 NUMBER OF SEQ ID NOS: 631
 SOFTWARE: PatentIn Ver. 2.0
 SEQ ID NO 49
 LENGTH: 2176
 TYPE: DNA
 ORGANISM: Homo sapiens
 US-09-822-830A-49

Alignment Scores:
 Pred. No.: 4.34e-155 Length: 2176
 Score: 1922.00 Matches: 338
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 9 Gaps: 0

US-10-791-017a-2_COPY_319_656 (1-338) x US-09-822-830A-49 (1-2176)

QY 1 G1YGYAARG1YGLYMETG1YSER1AG1YGLUARG1YGLYPHEANLYSPROGLYGLY 20
 DB 1207 GGAGGAGCGCGTGGAAATGGGCGAGCGTGGAGAGAGGTGGCTTCAATTAAGCGTGGGA 1148
 QY 21 PROMETASPG1UGLYPROAEPLEUASPLEUGLYPROFOVALASPPROAEPGLUASPSER 40
 DB 1147 CCCATGATGAAGAGCAGATCTTTGATCTGAGGCCCACTTGAATCAAGATGAAGACTCT 1088
 QY 41 ASAPSASER1A1ETRYVALING1YLEUASAPSASER1A1ETRYVALING1YLEUASAPS 60
 DB 1087 GACACAGTGCATTTAT 1028

QY 61 AspPhepHehYsgInCySgIyVAlValLyMeCAsnLySArGThrGlyGInPMeC11e 80
DB 1027 GACTTCTTAAAGCAGTGTGGGTTGTTAAAGTAAGAACTGGGCAACCCATGATC 968
QY 81 HisIIeTyLeuAspLySgIuThrGlyLySProLySgIyAspAlaThrValSerTyGlu 100
DB 967 CACATCTAACTGACAAAGAAACAGAAAGCCCAAGGCCATGCCACATGTCTTATGAA 908
QY 101 AspProProThraAlaLySAlaAlaValGluTrpPheAspGlyLySAspPheGInLySer 120
DB 907 GACCCACCCCACTGCAAGGCTGCCGTGAATGTTTATGGAAGAAAGATTTTCAAGGAGC 848
QY 121 LyLeuLySValSerLeuAlaArgLySlySProProMetAsnSerMetAlaGlyLyLeu 140
DB 847 AAACCTTAAGTCTCCCTGTGCTCCGAAAGACCTCCAAATGAACAGTATGCGGGGTGCTG 788
QY 141 ProProArgGluGlyAArgGlyMetProProProLeuArgGlyLyProGlyLyProGly 160
DB 787 CCAACCCGTGAAGGAGAGGCAATGCCACCACTCCGTGAAGGTCCAGAGGCCCAAG 728
QY 161 GlyProGlyLyProMetGlyAArgMetGlyLyAArgGlyLyAAspArgGlyLyPhePro 180
DB 727 GGTCTGGGGGAGCCCAATGGGTGCGCATGGAGGCCGTGAGAGATAGAGAGGCTTCCCT 668
QY 181 ProArgGlyLyProArgGlySerArgGlyAAspProSerGlyLyGlyAAsnValGlnHisArg 200
DB 667 CCAAGAGAACCCCGGGGTTCCTCCAGGGAGACCTCTGAGAGAGAAACGTCCAGACCGA 608
QY 201 AlaGlyAspTrpGlnCySProAsnProGlyCySgIyVAsnGlnAsnPheAlaTrpArgThr 220
DB 607 GCTGAGACCTGCGAGTCTCCCAATCCGGGTGTGGAACAGAACTTCCCTGGAGAAC 548
QY 221 GlnCyAsnGlnCySlySAlaProLySProGlnGlyPheLeuProProPhePro 240
DB 547 GAGTGCACACAGTGAAGGCCCAAGCCTGAAGGCTTCTCCCGCAACCTTTCCGCC 488
QY 241 ProGlyLyAAspArgGlyAArgGlyLyProGlyLyMetArgGlyLyAArgGlyLyLeu 260
DB 487 CCGGGGTGTATGTGTGCAAGAGTGGCCCTGTGTGGCAATGGGGAGAAAGGTGGCTC 428
QY 261 MetAspArgGlyLyProGlyLyMetPheArgGlyLyAArgGlyLyAAspArgGlyLy 280
DB 427 ATGATCTGTGTGTCTCCGATGATGTTCAGAGGTGGCCGTGTGAGCAAGAGGTGC 368
QY 281 PheArgGlyLyAArgGlyMetAspArgGlyLyPheGlyLyGlyAArgGlyLyPro 300
DB 367 TTCGGGTGTGGCCGGGCAATGACCAAGGTGGCTTGTGTGAGAGAAACAGGTGGCCCT 308
QY 301 GlyLyProProGlyLyProLeuMetGlnMetGlyLyAArgArgGlyLyAArgGlyLy 320
DB 307 GGGGGGCCCCCTTGACCTTGAATGGAACAGATGGAGAGAAAGAGAGGACGTGGAGGA 248
QY 321 ProGlyLyMetAspLySgIyGlnHisArgGlnGlnArgAspArgProTy 338
DB 247 CTTGAAAAATGATTAAGGCGAGACCTGTCAGAGCGCAGAGATCGGCCCTTAC 194

RESULT 3
US-09-880-107-3769
; Sequence 3769, Application US/09880107
; Patent No. US2002014281A1
; GENERAL INFORMATION:
; APPLICANT: Horne, Darci T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Scherf, Uwe
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
; FILE REFERENCE: 44921-5028-WO
; CURRENT APPLICATION NUMBER: US/09/880,107
; CURRENT FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/211,379
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/237,054

; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 3950
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3769
; LENGTH: 2390
; TYPE: DNA
; ORGANISM: Homo sapiens
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 X66899
US-09-880-107-3769

Alignment Scores:
Pred. No.: 4 73e-155 Length: 2390
Score: 1922.00 Matches: 338
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 9 Gaps: 0

US-10-791-017a-2_COPY_319_656 (1-338) x US-09-880-107-3769 (1-2390)

QY 1 GlyLyAArgGlyLyMetGlySerAlaGlyLyAArgGlyLyPheAsnLySProGlyLy 20
DB 998 GAGAGAGCGGTGGAATGGGCAAGCTGAGAGAGAGGTGCTTCAATAGCCGTGGA 1057
QY 21 ProMetAspGluGlyProAspLeuAspLeuGlyProProValAspProAspGlyAspSer 40
DB 1058 CCCATGATTAAGAGACAGATCTTGAATCTAGGCCCTCTGTATCCAGATGAAGACTCT 1117
QY 41 AspAsnSerAlaLeTyTrValGlnGlyLeuAsnAspSerValThrLeuAspLeuAla 60
DB 1118 GACACAGTGCATTAATTAATGACAGATTAATGACAGTGTACTAGATGATCTGGA 1177
QY 61 AspPhepHehYsgInCySgIyVAlValLyMeCAsnLySArGThrGlyGInPMeC11e 80
DB 1178 GACTTCTTAAAGCAGTGTGGGTTGTTAAAGTAAGAACTGGGCAACCATGATC 1237
QY 81 HisIIeTyLeuAspLySgIuThrGlyLySProLySgIyAspAlaThrValSerTyGlu 100
DB 1238 CACATCTAACTGACAAAGAAACAGAAAGCCCAAGGCCATGCCACATGTCTTATGAA 1297
QY 101 AspProProThraAlaLySAlaProLySProGlnGlyPheLeuProProPhePro 120
DB 1298 GACCCACCCCACTGCAAGGCTGCCGTGAATGTTGTAAGTGAAGATTTTCAAGGAGC 1357
QY 121 LyLeuLySValSerLeuAlaArgLySlySProProMetAsnSerMetAlaGlyLyLeu 140
DB 1358 AAACCTTAAGTCTCCCTTGCTGGAAGAACCTCCATAGAACATGCGGGGTGTG 1417
QY 141 ProProArgGluGlyAArgGlyMetProProProLeuArgGlyLyProGlyLyProGly 160
DB 1418 CCAACCCGTGAAGGAGAGGCAATGCCACCACTCCGTGAGAGTCCAGAGGCCCAAG 1477
QY 161 GlyProGlyLyProMetGlyAArgMetGlyLyAArgGlyLyAAspArgGlyLyPhePro 180
DB 1478 GGTCTGGGGGAGCCCAATGGGTGCGCATGGAGGCCGTGTGAGAAATGAGAGGCTTCCCT 1537
QY 181 ProArgGlyLyProArgGlySerArgGlyAAspProSerGlyLyGlyAAsnValGlnHisArg 200
DB 1538 CCAAGAGAACCCCGGGGTTCCTCCAGAGAACCCCTCTGAGAGAGAAACGTCCAGACCGA 1597
QY 201 AlaGlyAspTrpGlnCySProAsnProGlyCySgIyVAsnGlnAsnPheAlaTrpArgThr 220
DB 1598 GCTGAGACCTGCGAGTCCCAATCCGGGTGTGGAACCAAGAACTTCCGTGAGAGAA 1657
QY 221 GlnCyAsnGlnCySlySAlaProLySProGlnGlyPheLeuProProPhePro 240
DB 1658 GAGTGCACACAGTGAAGGCCCAAGACCTGAAGGCTTCTCCCGCAACCTTTCCGCC 1717
QY 241 ProGlyLyAAspArgGlyAArgGlyLyProGlyLyMetArgGlyLyAArgGlyLyLeu 260
DB 1718 CCGGGGTGTATGTGTGCAAGAGTGGCCCTGTGTGGAATCGGGGAGAAAGAGGTGGCTC 1777

QY 261 MetAspArgGlyValProGlyValMetPheArgGlyValArgGlyValAspArgGlyVal 280
DB 1778 ATGATCGTGTGTGTCCTGGGAGAAATGTTAGAGAGTGGCCGTGTGTGAGACAGAGGTGGC 1837
QY 281 PheArgGlyValArgGlyValMetAspArgGlyValPheGlyValValArgArgGlyValPro 300
DB 1838 TTCCTGTGTGTGGCCGGGAGATGAGACGAGGTGGCTTGTGTGTGAGAGAGAGAGAGTGGCCCT 1897
QY 301 GlyValProProGlyValProLeuMetGluGlnMetGlyValValArgArgGlyValArgGlyVal 320
DB 1898 GGGGGGGCCCCCTGAGCCTTTGATGTGAACAGATGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1957
QY 321 ProGlyValMetAspArgGlyValHisArgGlnGluArgArgAspArgProTyr 338
DB 1958 CCTGAGAAATGATTAAGGAGAGACCTGACAGAGCCGAGAGATGGCCCTTAC 2011

RESULT 4
US-09-960-706-1081
; Sequence 1081, Application US/09960706
; Publication No. US20030134280A1
; GENERAL INFORMATION:
; APPLICANT: Munger, William E.
; TITLE OF INVENTION: Identifying Drugs for and Diagnosis of Benign Prostatic Hyperplasia
; FILE REFERENCE: 44921-5029-01US
; CURRENT APPLICATION NUMBER: US/09/960,706
; CURRENT FILING DATE: 2001-09-24
; PRIOR APPLICATION NUMBER: 60/223,323
; PRIOR FILING DATE: 2000-08-07
; PRIOR APPLICATION NUMBER: 09/873,319
; PRIOR FILING DATE: 2001-06-05
; NUMBER OF SEQ ID NOS: 1124
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 1081
; LENGTH: 2390
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20030134280A1 X66899
US-09-960-706-1081

Alignment Scores:
Pred. No.: 4,73e-155 Length: 2390
Score: 1922.00 Matches: 338
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 10 Gaps: 0

US-10-791-017a-2_copy_319_656 (1-338) x US-09-960-706-1081 (1-2390)

QY 1 GlyValArgGlyValMetGlySerAlaGlyValArgGlyValPheAlaValProGlyVal 20
DB 998 GAGAGAGCGCGGTGAAATGGGACGCGTGGAGAGCGGAGGTGGCTTCAATAGCCGTGGTGA 1057
QY 21 PrometAspGluGlyProAspLeuAspLeuGlyProProValAspProAspGluAspSer 40
DB 1058 CCCATGATGAAGACCAAGATCTTGAATGAGCCCTCTTGAATGCCAGATGAAGACTCT 1117
QY 41 AspAsnSerAlaIleTyrValGlnGlyLeuAsnAspSerValThrLeuAspAspLeuAla 60
DB 1118 GACACAGTGCATTTATGTAACAGATTAATATACAGTGTGACTCTATGATGATCGCA 1177
QY 61 AspPhePheLeuGlnGlyValValValMetAsnValArgThrGlyGlnProMetIle 80
DB 1178 GACTTCTTTAAGCAGTGTGGGTGTTAATGATGAACAGAGAACTGGGCAACCATGATC 1237
QY 81 HisIleTyrLeuAspArgGluThrGlyValProGlyValAspAlaThrValSerTyrGlu 100
DB 1238 CACATCTACCTGACAGAGAAACAGAAAGCCCAAGCCCAATGCCAAGTGTCTATGAA 1297
QY 101 AspProThrAlaValAlaValAlaValGluTyrPheAspGlyValAspPheGlnGlySer 120

DB 1298 GACCCACCCACTGCCAAGGCTCCGTGATGATGAGAAATTTTCAGAGGAGC 1357
QY 121 LysLeuLysValSerLeuAlaArgLysValProPrometAsnSerMetArgGlyValLeu 140
DB 1358 AAACCTTAAGTCTCCCTTGTCTGGAAGAGCCCTCCATATACAGTATGGGGGTGTCTG 1417
QY 141 ProProArgGlyGlyValArgGlyValMetProProProLeuArgGlyValProGlyValProGlyVal 160
DB 1418 CACACCCCGTGAAGGAGAGAGGATGCCACCACTCTCCGTGAGAGTCAAGAGGCCACAGAG 1477
QY 161 GlyProGlyValProMetGlyValArgMetGlyValArgGlyValAspArgGlyValPhePro 180
DB 1478 GGTCTGGGGAGCCCATGTGTGTCATGTGGAGCCCGTGGAGAGATGAGAGAGGCTTCCCT 1537
QY 181 ProArgGlyProArgGlySerArgGlyAsnProSerGlyValGlyValAsnValGlnHisArg 200
DB 1538 CCAGAGAGACCCCGGGGTTCCTGAGAGAAACCCCTCTGAGAGAGAAACGTCAGACCGA 1597
QY 201 AlaGlyAspTyrGlnCysProAsnProGlyCysGlyAsnGlnAsnPheAlaTyrArgThr 220
DB 1598 GCTGAGACTGACAGTGTCTCCATCCGGTTGTGAAACAGAACTTCCCTGAGAGAA 1657
QY 221 GluCysAsnGlnCysValAlaProLysProGlnGlyPheLeuProProPheProPro 240
DB 1658 GAGTGCAACAGTGTAAAGCCCAAGAGCTTCTCCGCGCACCTTCCGCC 1717
QY 241 ProGlyValAspArgGlyValArgGlyValProGlyValMetArgGlyValArgGlyValLeu 260
DB 1718 CCGGGTGTGATCGTGGCAGAGGTGCGCTGTGTCATCGGGAGAGAGAGGTGCTC 1777
QY 261 MetAspArgGlyValProGlyValMetPheArgGlyValArgGlyValValAspArgGlyVal 280
DB 1778 ATGATCGTGTGTGTCCTGGGAGAAATGTTCAAGAGTGGCCGTGTGAGACAGAGGTGGC 1837
QY 281 PheArgGlyValArgGlyValMetAspArgGlyValPheGlyValValArgArgGlyValPro 300
DB 1838 TTCCTGTGTGTGGCCGGGAGATGAGACGAGGTGGCTTGTGTGAGAGAGAGAGGTGGCCCT 1897
QY 301 GlyValProProGlyValProLeuMetGluGlnMetGlyValValArgArgGlyValArgGlyVal 320
DB 1898 GGGGGGGCCCCCTGAGCCTTTGATGTGAACAGATGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1957

RESULT 5
US-09-873-319-717
; Sequence 717, Application US/09873319A
; Publication No. US20030134324A1
; GENERAL INFORMATION:
; APPLICANT: Munger, William E.
; APPLICANT: Kulkarni, Prakash
; APPLICANT: Getzenberg, Robert H.
; APPLICANT: Waga, Iwao
; APPLICANT: Yamamoto, Jun
; TITLE OF INVENTION: Identifying Drugs for and Diagnosis of Benign Prostatic
; FILE REFERENCE: 44921-5029-US
; CURRENT APPLICATION NUMBER: US/09/873,319A
; EARLIER FILING DATE: 2001-06-05
; EARLIER APPLICATION NUMBER: US 60/223,323
; NUMBER OF SEQ ID NOS: 755
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 717
; LENGTH: 2390
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20030134324A1 X66899
US-09-873-319-717

Alignment Scores:

Pred. No.: 4,73e-155 Length: 2390
Score: 1922.00 Matches: 338
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 10 Gaps: 0

US-10-791-017a-2_COPY_319_656 (1-338) x US-09-873-319-717 (1-2390)

QY 1 G1YGLYARGLYGLYMETGLYSERIALAGLYGLUARGLYGLYPHEANLYSPROGLYGLY 20
DB 998 GGAGGAGCGCGGTGAATGGGCAAGGCTGGAGAGGAGAGTGGCTTCATTAAGCTGGGCA 1057
QY 21 PROMETAPGLUGLYPROAPLEUAPLEUGLYPROPROVALAPPROAPGLUAPPSER 40
DB 1058 CCCATGGATGAGGACAGATCTTGATCTAGCCCTCCTGTAGATCCAGATGAAAGACTCT 1117
QY 41 APPANSERIALIETRYVALGNGLYLEUASNSERVALTHREUASAPLEUALA 60
DB 1118 GACAAACAGTGCATTTATGTACAGAGATTAAATGACGTGTGACTTGAATGATCTGGCA 1177
QY 61 APPPHEHELHLYSGLYVALVALYMETASNLVSARGTTHRGYGLYINPROMETILE 80
DB 1178 GACTTCTTTAAGCAGTGTGGGTTTGAATGACAGAGAACTGGGCAACCATGATC 1237
QY 81 HIELIETRYLEUASPLYSGLUTHRGLYLYSPROLYSGLYASPLATHVALSERTRYGLU 100
DB 1238 CACATCTACTGACAGAGAAACAGAAAGCCCAAGGCGCATGACAGTGTCTTATGAA 1297
QY 101 APPPROPTHRALIYSAIAAVALGLUTRPHENAPGLYLYSAPPHEGLNGLYSER 120
DB 1298 GACCCCACTGACCTGCAAGGCTGCGTGAATGTTTATGAGAAAGATTTTCAAGGGAGC 1357
QY 121 LYSLEULYSVALSERLEUALAARGLYLYSPROPROMETASNSERMETARGLYGLYLEU 140
DB 1358 AAACCTTAAGTCTCCCTTCTCGGAGAGAACCTTCATGAACTATGCGGGGTGCTG 1417
QY 141 PROPROARGGLUGLYARGLYMETPROPROLEUARGGLYGLYPROGLYGLYPROGLY 160
DB 1418 CCAACCCGAGGAGCAAGGAGCATGCAACCACTCCTGGAGGCTCAGAGAGGCCCAAGGA 1477
QY 161 GLYPROGLYGLYPROMETGLYARGMETGLYGLYARGGLYGLYASPARGLYGLYPHEPRO 180
DB 1478 GGTCTGGGGGAGCCCATGGGTGCGATGGAGGCGGTGAGAGAGATGAGAGAGGCTTCCCT 1537
QY 181 PROARGGLYPROARGGLYSERARGGLYASNSERSEGLYGLYGLYASNSVALGNIHLEARG 200
DB 1538 CCAAGAGAGACCCCGGGGTTCCTCGAGAGAACCTCTTGGAGAGAGAACTGCAACCCGA 1597
QY 201 ALIAGLYASPTIRPGLNCYSPROASNPROGLYCYSGLYANGLNANPHEALATIRPARGTHR 220
DB 1598 GCTGGAGACTGGCAGTGTCCCAATCCGGGTTGTGGAACCAAGAACTTCGCTGGAGAA 1657
QY 221 GLUCYABANGLYCYBLYSALAPROLYSPROGLUGLYPHELEUPROPROPHETPRO 240
DB 1658 GAATGCAACCACTGAAGGCCCAAMAGCTGAAAGGCTTCTCCGCAACCTTCCGCCCC 1717
QY 241 PROGLYGLYASPARGLYARGGLYGLYPROGLYGLYMETARGLYGLYARGGLYGLYLEU 260
DB 1718 CCGGGGTGTATCTGGCAAGGAGGTCCTGTGTGCAATGCGGGAGAGAAAGGTGACCTC 1777
QY 261 METASPARGLYGLYPROGLYGLYMETPHEARGGLYGLYARGGLYGLYASPARGLYGLY 280
DB 1778 ATGAGATGTGTGTGCTCCGGTGGAAATGTTCAAGAGTGGCCGTGTGAGACAGAGGTGGC 1837
QY 281 PHEARGGLYGLYARGGLYMETASPARGLYGLYPHEGLYGLYGLYARGGLYGLYPRO 300
DB 1838 TTCCTGTGTGGCGGGGAGATGAGCAAGGTGCTTTGTGTGAGAGAAACAGAGGTGGCCCT 1897
QY 301 GLYGLYPROPROGLYPROLEUMETGLNMEGLYGLYARGARGGLYGLYARGGLYGLY 320
DB 1898 GGGGGGCCCCCTGGACTTTGATGGAACAGATGGAGAGAAAGAGAGAGACGTGGAGGA 1957

QY 321 PROGLYLYMETASPLYSGLYGLUHIARGGLNGLUARGARGASPARGTPT 338
DB 1958 CCTGGAAAAATGATTAAGGCGAGCACCGTCAAGAGGCGAGAGATCGGCCCTTAC 2011

RESULT 6

US-10-791-017a-1
; Sequence 1, Application US/10791017A
; Publication No. US20040197827A1
; GENERAL INFORMATION:
; APPLICANT: JENAPHARM GmbH & Co. KG
; TITLE OF INVENTION: Methods for Determining Hormonal Effects of Substances
; FILE REFERENCE: Pat 3684/11
; CURRENT APPLICATION NUMBER: US/10/791, 017A
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 2390
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURES:
; NAME/KEY: CDS
; LOCATION: (44)..(2011)
; OTHER INFORMATION: EMS
US-10-791-017a-1

Alignment Scores:

Pred. No.: 4,73e-155 Length: 2390
Score: 1922.00 Matches: 338
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 18 Gaps: 0

US-10-791-017a-2_COPY_319_656 (1-338) x US-10-791-017a-1 (1-2390)

QY 1 G1YGLYARGLYGLYMETGLYSERIALAGLYGLUARGLYGLYPHEANLYSPROGLYGLY 20
DB 998 GGAGGAGCGCGGTGAATGGGCAAGGCTGGAGAGGAGAGTGGCTTCATTAAGCTGGGCA 1057
QY 21 PROMETAPGLUGLYPROAPLEUAPLEUGLYPROPROVALAPPROAPGLUAPPSER 40
DB 1058 CCCATGGATGAGGACAGATCTTGATCTAGCCCTCCTGTAGATCCAGATGAAAGACTCT 1117
QY 41 APPANSERIALIETRYVALGNGLYLEUASNSERVALTHREUASAPLEUALA 60
DB 1118 GACAAACAGTGCATTTATGTACAGAGATTAAATGACAGTGTGACTTGAATGATCTGGCA 1177
QY 61 APPPHEHELHLYSGLYVALVALYMETASNLVSARGTTHRGYGLYINPROMETILE 80
DB 1178 GACTTCTTTAAGCAGTGTGGGTTTGAATGACAGAGAACTGGGCAACCATGATC 1237
QY 81 HIELIETRYLEUASPLYSGLUTHRGLYLYSPROLYSGLYASPLATHVALSERTRYGLU 100
DB 1238 CACATCTACTGACAGAGAAACAGAAAGCCCAAGGCGCATGACAGTGTCTTATGAA 1297
QY 101 APPPROPTHRALIYSAIAAVALGLUTRPHENAPGLYLYSAPPHEGLNGLYSER 120
DB 1298 GACCCCACTGACCTGCAAGGCTGCGTGAATGTTTGAAGAGATTTTCAAGGGAGC 1357
QY 121 LYSLEULYSVALSERLEUALAARGLYLYSPROPROMETASNSERMETARGLYGLYLEU 140
DB 1358 AAACCTTAAGTCTCCCTTCTCGGAGAGAACCTTCATGAACTATGCGGGGTGCTG 1417
QY 141 PROPROARGGLUGLYARGLYMETPROPROLEUARGGLYGLYPROGLYGLYPROGLY 160
DB 1418 CCAACCCGAGGAGCAAGGAGCATGCAACCACTCCTGTGAGAGTCCAGAGAGGCCCAAGGA 1477
QY 161 GLYPROGLYGLYPROMETGLYARGMETGLYGLYARGGLYGLYASPARGLYGLYPHEPRO 180
DB 1478 GGTCTGGGGGAGCCCATGGGTGCGATGGAGGCGGTGAGAGAGATGAGAGAGGCTTCCCT 1537

QY 181 ProArg1yProArg1ySerArg1yAsnProSerG1yG1yG1yAsnValG1nH1sArg 200
DB 1538 CCAAGAGACCCCGGGGTTCCAGAGGAACTCTGAGAGAGAAAGTCCAGACCGA 1597
QY 201 AlaG1yAspTrpG1nCysProAsnProG1yCysG1yAsnG1nAsnPheA1aTrpArgThr 220
DB 1598 GCTGAGACTGGAGTGTCTCCCAATCCGGTTGTGGAACAGAACTTCGCTGGAGACA 1657
QY 221 GluCyAsnGlnCysGlyAlaProLysProG1nG1yPheLeuProProPheProPro 240
DB 1658 GAGTGCACACAGGTAAAGGCCCAAGAGGCTTCTCCCGCACCTTCGCGCC 1717
QY 241 ProG1yG1yAspArg1yArg1yG1yPheG1yG1yMetArg1yG1yArg1yG1yLeu 260
DB 1718 CCGGTTGTATCTGTGCAAGGTGCTGTGTGCTGTGTGCTGTGTGCTGTGTGCT 1777
QY 261 MetAspArg1yG1yProG1yG1yMetPheArg1yG1yArg1yG1yAspArg1yG1y 280
DB 1778 ATGGATCTGTGTGTCTCCGGTGGAAATGTCAGAGTGTGCTGTGTGAGACAGAGTGC 1837
QY 281 PheArg1yG1yArg1yMetAspArg1yG1yPheG1yG1yG1yArg1yG1yPro 300
DB 1838 TTCGGTGTGTGCGGGGCAAGACCGAGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1897
QY 301 G1yG1yProProG1yProLeuMetG1nMetG1yG1yArg1yG1yArg1yG1y 320
DB 1898 GGGGGGCGCTTGTGCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1957
QY 321 ProG1yLysMetAspLysG1yG1yG1yG1yG1yG1yG1yG1yG1yG1y 338
DB 1958 CTGGAATAATGATTAAGCGAGCACCTGAGAGCGAGATCGGCTTAC 2011

RESULT 7

US-09-822-830A-410/C
; Sequence 410, Application US/09822830A
; Patent No. US20020142952A1
; GENERAL INFORMATION:
; APPLICANT: Genetics Institute, Inc.
; APPLICANT: Mong, Gordon G.
; APPLICANT: Clark, Hilary
; APPLICANT: Fechtel, Kim
; APPLICANT: Agostino, Michael J.
; APPLICANT: Howes, Steven H.
; APPLICANT: Resnick, Richard J.
; APPLICANT: Gulukota, Kamalakari
; APPLICANT: Graham, James R.
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING NOVEL SECRETED PROTEINS
; FILE REFERENCE: GIN 6402
; CURRENT APPLICATION NUMBER: US/09/822,830A
; PRIOR FILING DATE: 2001-03-29
; PRIOR APPLICATION NUMBER: 60/195,604
; NUMBER OF SEQ ID NOS: 631
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 410
; LENGTH: 2273
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-822-830A-410

Alignment Scores:

Prod. No.: 1,13e-142 Length: 2273
Score: 1777.00 Matches: 319
Percent Similarity: 94.38% Conservative: 0
Best Local Similarity: 94.38% Mismatches: 1
Query Match: 92.46% Indels: 18
DB: 9 Gaps: 2

US-10-791-017a-2_COPY_319_656 (1-338) x US-09-822-830A-410 (1-2273)

QY 1 G1yG1yArg1yG1yMetG1ySerAlaG1yG1yG1yG1yPheAsnLysProG1yG1y 20
DB 1312 GGAAGACCGCGTGAATAGGCGC---GCTGAGAGAGAGAGTGTCTTCAATTAAGCCCTGCTGGA 1256

QY 21 ProMetAspGluG1yProAspLeuAspLeuG1yProProValAspProAspG1uAspSer 40
DB 1255 CCCATGATGAAAGAACCAATCTTGAATCTAGGCCACCTGTAAATCAAGTAAAGACTCT 1196
QY 41 AspAsnSerAlaLeuTyValGlnG1yLeuAsnAspSerValThrLeuAspLeuAla 60
DB 1195 GACCAAGTGCATTTATATATCAAGAGTTAAATGACAGTGTGACTCTAATATGATCTGGCA 1136
QY 61 AspPhePheLysG1nCysG1yValValLysMetAsnLysArgThrG1yG1nProMetIle 80
DB 1135 GACTTCTTTAGACAGTGTGGGTGTGTTAAGATGAACAGAGAACTGGCAACCCAGATC 1076
QY 81 HisLeuTyLeuAspLysG1uThrG1yLysProLysG1yAspAlaThrValSerTyGlu 100
DB 1075 CACATTAACCTTGACAGAAACAGAAAGCCCAAGGGGATGCCACAGTCTCTATGAA 1016
QY 101 AspProProThrAlaLysAlaAlaValG1uTrpPheAspG1yLysAspPheGlnG1ySer 120
DB 1015 GACCCACCCACATGCGCAAGGCTGCGGTGAATGTTTGTGGAAAGATTTTCAAGGAGC 956
QY 121 LysLeuLysValSerLeuAlaArgLysLysProProMetAsnSerMetArgG1yG1yLeu 140
DB 955 AAACCTTAAGTCTCCCTTGTCTCGGAAGAGCTCCATACAGTATGCGGGTGTGTCTG 896
QY 141 ProProArgG1nG1yArg1yMetProProProLeuArg1yG1yProG1yG1yProG1y 160
DB 895 CACCCCGTGAAGGACAGAGCAGTGCACCACTCGTGAAGGTCAGAGAGCCAGAGCA 836
QY 161 G1yProG1yG1yProMetG1yArgMetG1yG1yArg1yG1yAspArg1yG1yPhePro 180
DB 835 GGTCTTGGGGGACCCATGGGTGCGCATGGAGGCGGTGAGAGATAGAGAGGCTTCCCT 776
QY 181 ProArgG1yProArgG1ySerArg1yAsnProSerG1yG1yG1yAsnValG1nH1sArg 200
DB 775 CCAAGAGACCCCGGGGTTCCAGAGGAACTCTGTGAGAGAGAAACGTCAGACCGA 716
QY 201 AlaG1yAspTrpG1nCysProAsnProG1yCysG1yAsnG1nAsnPheA1aTrpArgThr 220
DB 715 GCTGAGACTGTCAGAGTGTCCCAATCCGGTTGTGGAACAGAACTTCGCTGAGAGACA 656
QY 221 GluCyAsnGlnCysGlyAlaProLysProG1nG1yPheLeuProProPheProPro 240
DB 655 GAGTGCACACAGT----- 641
QY 241 ProG1yLysMetAspLysG1yG1yG1yG1yG1yG1yG1yG1yG1yG1y 260
DB 640 -----GTTGATCTGTGCGAGAGGTGCTCTGTGTGTCATGTGGGAGAGAGAGTGTGCTC 587
QY 261 MetAspArgG1yG1yProG1yG1yMetPheArgG1yG1yArgG1yG1yAspArgG1yG1y 280
DB 586 ATGGATCTGTGTGTCTCCGGTGAATGTTCAAGGTGGCCGTGTGTGAGACAGAGTGTGC 527
QY 281 PheArgG1yG1yArg1yMetAspArgG1yG1yPheG1yG1yG1yArgArgG1yG1yPro 300
DB 526 TTCGGTGTGTGCGGGGCAATGACAGAGGTGCTTTGTGTGAGAAAGACAGAGTGTGCTC 467
QY 301 G1yG1yProProG1yProLeuMetG1nMetG1yG1yArgArgG1yG1yArgG1yG1y 320
DB 466 GGGGGGCGCTTGTGACCTTGTATGGAACAGATGGAGAGAAAGAGAGAGAGACCTGTGAGGA 407
QY 321 ProG1yLysMetAspLysG1yG1yG1yG1yG1yG1yG1yG1yG1yG1y 338
DB 406 CCGGAATAATGATTAAGCGAGCACCTGAGAGCGAGAGATCGGCTTAC 353

RESULT 8

US-10-198-846-9847
; Sequence 9847, Application US/10198846
; Publication No. US2003009974A1
; GENERAL INFORMATION:
; APPLICANT: Lillie, James
; APPLICANT: Xu, Yongyao
; APPLICANT: Wang, Youzhen

```
/ APPLICANT: Steinmann, Kathleen
/ TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
/ TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND...
/ FILE OF INVENTION: THERAPY OF BREAST CANCER
/ FILE REFERENCE: MRI-049
/ CURRENT APPLICATION NUMBER: US/10/198,846
/ CURRENT FILING DATE: 2002-07-18
/ PRIOR APPLICATION NUMBER: 60/306,220
/ PRIOR FILING DATE: 2001-07-18
/ NUMBER OF SEQ ID NOS: 14084
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 9847
/ LENGTH: 2299
/ TYPE: DNA
/ ORGANISM: Homo sapiens
US-10-198-846-9847

Alignment Scores:
Pred. No.: 1,05e-91 Length: 2299
Score: 1181.00 Matches: 222
Percent Similarity: 66.07% Conservative: 0
Best Local Similarity: 66.07% Mismatches: 1
Query Match: 61.45% Indels: 114
DB: 14 Gaps: 1

US-10-791-017a-2_COPY_319_656 (1-338) x US-10-198-846-9847 (1-2299)

QY 20 G1YPromerAapgluglYProaapleuapleuglYProbovalaapProaapgluap 39
DB 1072 GGAACCATGTAGTAGAGACCAAGATCTTGATCTAGGCCCATCTAGATCCAGATGAAGAC 1131
QY 40 SerAapSerAalietYrvalGinglYleuabnapSerValThrlaapAapleu 59
DB 1132 TCTGACAAAGTCAATTATGTACAAAGATTAATACAGTGTACTTATGATGATCTG 1191
QY 60 AlAapPhePheYsglYValValIyemecAmlYsaPThrglYglInProMet 79
DB 1192 GCAGACTTCTTTAGAGCTGTGGGTTGTATAGTAAACAGAGACTGGGCACACCATG 1251
QY 80 lIhIaIeTyrleuaplyglYlurthglYlyeProlYsglYsaPalathrValSerTyr 99
DB 1252 ATCCACATCTACCTGCGAAGAAACAGAAAGCCCAAGGCGATGCCACAGTGTCTAT 1311
QY 100 GluAapProProthralalyValaIaValaGlUtrPheaspglYlysaPheInglY 119
DB 1312 GAAGACCCACCCCTGCGAAGGCTGCGGTGAATGTTGATGGGAAAGATTTTCAAGCG 1371
QY 120 SerlyleuYValSerleuAlaYglYlyeProlYsaPheAapSerMetArgglY 139
DB 1372 AGCAACTTAAAGTCTCTGCTCGAAGAAAGCTTCAATGAACAGTATGCGGGTGT 1431
QY 140 leuPProAapgluglYarGlyMetProProProlYsaPheAapglYglYProglY 159
DB 1432 CTGCCACCCCGTGAAGGCGAAGGATGCGACCACTCCGTGAGGTCCAGAGGCGCA 1491
QY 160 GlYglYProglYglYProMetGlYarGlyMetGlYglYarGlyglYsaPheAapglY 179
DB 1492 GAGAGTCTCGGGGAGACCATGAGGTGCGATGAGAGGCGGTGAGAGATGAAGAGGCTTC 1551
QY 180 ProProAapglYProAapglYserAapglYsaPProSerGlYglYglYsaPValGlnHs 199
DB 1552 CTTCCAGAGGAGACCCCGGGGTTCCCGAGGAGAACCCCTCTGAGAGGAGAACGTCAGAC 1611
QY 200 ArgAlaGlYsaPTrpInCYsaProaapProglY----- 210
DB 1612 CGAGCTGAGAGACTGGAGGTGCCAATCCGTA-TGTACTTGTCTGGCAAAATTGATACC 1670
QY 210 ----- 210
DB 1671 TACGAGTGAAGCAACCTTCCCTCAACCCCATCTAGAGTGAATGCTCTGTCTA 1730
QY 210 ----- 210
```

```
DB 1731 GAGAAACAGATGATGACCTTGATGGGTAGGGAACCTAGACACCATGACTGAGAC 1790
QY 210 ----- 210
DB 1791 GCTTACAGACCTTCTGAGATGATTTGACCTGTCTGTGGTGAATGCTGCTGAGGC 1850
QY 210 ----- 210
DB 1851 TGTGCTTAAGATGGGTGATGATGATCTCTGATGATGATGATGATGATGATGATGAT 1910
QY 210 ----- 210
DB 1911 CACCACTTCTGTTTATCTTCTTATCTTCAATGATGATTTCTGATGATGAT 1970
QY 211 -----CysglYanGlnaapPheAlaTrpAargThrglYcyAanGlnCYsly 226
DB 1971 GTATGAGAGGTTGTGGAACAGAACTTGTGCTGAGAGAACAGATGACAGTGTAAAG 2030
QY 227 AlAProlyProgluglYpHeleuProProPheProProglY 242
DB 2031 GCCCAAGCTGAAGGCTTCTCCGCGCACCTTCCGCCCGGGGT 2078

RESULT 9
US-09-918-995-27690
/ Sequence 27690, Application US/09918995
/ Publication No. US20030073623A1
/ GENERAL INFORMATION:
/ APPLICANT: Hyseq, Inc.
/ TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
/ TITLE OF INVENTION: FROM VARIOUS CDNA LIBRARIES
/ FILE REFERENCE: 20411-756
/ CURRENT APPLICATION NUMBER: US/09/918,995
/ CURRENT FILING DATE: 2001-07-30
/ PRIOR APPLICATION NUMBER: US/09/235,076
/ PRIOR FILING DATE: 1999-01-20
/ NUMBER OF SEQ ID NOS: 38054
/ SOFTWARE: FastSeq for Windows Version 3.0
/ SEQ ID NO 27690
/ LENGTH: 550
/ TYPE: DNA
/ ORGANISM: Homo sapiens
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (1)..(550)
/ OTHER INFORMATION: n = A,T,C or G
US-09-918-995-27690

Alignment Scores:
Pred. No.: 1,86e-73 Length: 550
Score: 961.00 Matches: 166
Percent Similarity: 97.65% Conservative: 0
Best Local Similarity: 97.65% Mismatches: 4
Query Match: 50.00% Indels: 0
DB: 10 Gaps: 0

US-10-791-017a-2_COPY_319_656 (1-338) x US-09-918-995-27690 (1-550)

QY 119 G1YserlyleuYValSerleuAlaYglYlyeProlYsaPheAapSerMetArgglY 138
DB 40 GGAAGCAACTTAAAGTCTCTGCTCGAAGAAAGCTTCAATGAACAGTATGCGGTGT 99
QY 139 G1YleuPProAapgluglYarGlyMetProProProlYsaPheAapglYglYProglY 158
DB 100 GGTCTGCCACCCCGTGAAGGCGAAGGATGCGACCACTCCGTGAGGTCCAGAGGCG 159
QY 159 ProglYglYProglYglYProMetGlYarGlyMetGlYglYarGlyglYsaPheAapglY 178
DB 160 CGAGAGTCTCGGGGAGACCATGAGGTGCGATGAGAGGCGGTGAGAGATGAAGAGGCG 219
QY 179 PheProProAapglYProAapglYserAapglYsaPProSerGlYglYglYsaPValGln 198
DB 220 TTCCCTCAAGAGAGACCCCGGGGTTCCCGAGGAGAACCCCTCTGAGAGGAGAAAGCTCCAG 279
```

Oy	199	115ATGAG1AG1YA8PTTbG1nCYbProA8nProG1CYbG1YA8nG1nA8nPhealATP	218
Db	280	CACCA8GTG8AGACTG8CAGTGTCCCAATCCGGGTGTGTG8AAACCA8AACTTCG8CTG8	339
Oy	219	ATGThrg1nCYbA8nG1nCYbAla1aP8oLYbProG1nG1YbPhe1e8nP8oP8oPhe	238
Db	340	AGAAACAG8GTG8AACCA8CAGT8TAAG8GCCCA8AA8CCTGA8AG8CTTCCTCC8GC8CA8CCTTT	399
Oy	239	ProP8oProG1nG1YA8P8aT8rG1YA8T8rG1YbG1YProG1YbG1YMe8TA8rG1YbAT8rG1Y	258
Db	400	CCGCCCCCG8GTGTGTGTCTGT8GCAG8GTGT8G8CCTGTGTGT8CAT8GG8AG8AA8A8GT	459
Oy	259	G1YLeuMe8TA8P8aT8rG1YbG1YProG1YbG1YMe8T8Phe8rG1YbG1YA8T8rG1YA8P8aT8r	278
Db	460	GGCCTTCAT8GT8AT8CTGT8TGT8TCCCG8GT8GA8AT8TTCAT8GT8T8G8C8GT8GT8G8ACAG	519
Oy	279	G1YG1YbPhe8rG1YbG1YA8T8rG1YbMe8AP	288
Db	520	GGT8G8CTTC8GT8GT8T8G8CC8G8G8G8CA8TG8AC	549

```

RESULT 10
US-10-755-889-649
; Sequence 649, Application US/10755889
; Publication No. US20040171823A1
; GENERAL INFORMATION:
; APPLICANT: Bi-Iscol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES ASSOCIATED WITH THE NF-KB
; TITLE OF INVENTION: PATHWAY
; FILE REFERENCE: D0284 NP
; CURRENT APPLICATION NUMBER: US/10/755,889
; CURRENT FILING DATE: 2004-01-13
; PRIOR APPLICATION NUMBER: U.S. 60/440,068
; PRIOR FILING DATE: 2003-01-14
; PRIOR APPLICATION NUMBER: U.S. 60/469,757
; PRIOR FILING DATE: 2003-05-12
; NUMBER OF SEQ ID NOS: 823
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 649
; LENGTH: 1822
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-755-889-649

```

Alignment Scores:	
Pred. No.:	6,92e-66
Score:	876.00
Percent Similarity:	63.37%
Best Local Similarity:	52.91%
Query Match:	45,688
DB:	18
Length:	1822
Matches:	182
Conservative:	36
Mismatches:	54
Indels:	72
Gaps:	13

US-10-791-017A-2_COPY_319_656 (1-338) X US-10-755-889-649 (1-1822)

QY	1	GIYGIYAAGYGIYGIYMeGIYSerIaGIYLuIaYGIYGIYGIYPheAsnIYProGIYGIY	20
Db	814	GGAGGCAGAGGTGGCATGGGCGGAAGT---GACCGTGGTGGCTTCAAATTAATTTGGTGGC	870
QY	21	ProMetAspGIYGIYProAspLeuAspLeuGIYProValAspProAspGIYAspSer	40
Db	871	CCTCGGAGCCAGAGATTCACGTGCATAC-----TCCGAACAGAGAAATATTC	915
QY	41	AspAsnSerAlaIleTyrValGIYGIYLeuAsnAspSerValThrLeuAspAspLeuAla	60
Db	916	GACAAACAAACATCTTTGTGGCAAGCGCTGGGTGAGAAATGTTACATTTGAGCTGTGGCT	975
QY	61	AspPhePheIYGIYCyGIYValValIYsMetAsnIYsArgThrGIYGIYProMetIle	80
Db	976	GATTACTTCAGCAAGATTTGTATTATTATTAAGCAAAAGAAAGCGGACAGCCCATGATT	1035
QY	81	HisIleIYrIeAspIYsGIYThrGIYIYsProIYsGIYAspAlaIThrValSerIYrGIY	100
Db	1036	AATTTGTACACAGACAGCAAACTGGCAAGCTGAAGGAGAGGCAACGCTCTCTTTTGAAT	1095

Qy	101	AspProChtlralAlaValAlaAlaValGluTtrpPheAspGlyLysAspPheGlnGlySer	120
	
Db	1096	GAACCACCTTCAGCTAAAGCAGCATATTGACTGGTTTCATGCTAAAGAATTCCTCGGAAT	11555
Qy	121	LyseLeuLysValSerLeuAlaArgLysLysProProMetCenSerMetArgGlyLysLeu	140
	
Db	1156	CCATTCAAGGCTCATTTGCTACTCGCGGGGACAGACTTTAAT-----CGGGGCTGGTGC	12099
Qy	141	ProProArgGluGlyArgGlyMetProProProLeuArgGlyGlyProGlyLysProGly	160
Db	1210	AATGGTGGTGGAGGCCAGGG-----CGAGAGAGACCATGGCCGATGGGA	1254
Qy	161	GlyProGlyGlyProMetGlyArgMetGlyLysArgGlyLysAspArgGlyLysLysPhePro	180
Db	1255	GGCATATGAGAGTGGT-----GGAGAGTGGTGGTGGCCAGAGGAGATTTCC	1302
Qy	181	ProArgGlyProArgGlySerArgGlyAsnProSerGlyGlyGlyAsnValGlnHisArg	200
Db	1303	AGTGGAGGT-----GGTGGCGGTGGAGGACAGCAGCGGA	1335

QY	201	AlaGlyAspTrpGlnCysProAsnProGlyCysGlyAsnGlnPheAlaTrpArgThr	220
Db	1336	GCTGTGCTCGGAAGTGTCTTAATCCACCTGTGAATATGAATTAATCTTCTTGAGGAAAT	1395
QY	221	GIuCyAsnGlnCysIleAlaProIysProGluGlyPheLeuProProPheProPro	240
Db	1396	GAATGCACACCGTGTAAAGGCCCTTAAACCAAGATGC-----	1431
QY	241	ProGlyGlyAspArgGlyArgGlyGlyProGlyGlyIleMetArg---GlyGlyArgGlyGly	259
Db	1432	CCAGGA-----GGGGGACCAAGTGGCTCTCACATGGGGGGTAACTACGGG	1476
QY	260	LeuMetAspArgGlyGlyProGlyGlyIleMetPheArgGly-----GlyArgGlyGly	276
Db	1477	GATGATCGTGTGTGGCAGAGGAGCTATGATGAGAGCGGCTACCGGGGCCGGCGGG	1536
QY	277	AspArgGlyGlyPheArgGlyGlyArg---GlyMetAspArgGlyGlyPheGlyGlyGly	295
Db	1537	GACCGTGAAGCTTCCAGAGGGGCCGGTGTGGAGACAGAGTGGCTTT-----	1587
QY	296	ArgArgGlyGlyProGlyGlyProProGlyProLeuMetGluInMetGlyIleArgArg	315
Db	1587	-----	1587
QY	316	GlyGlyArgGlyGlyProGlyGlyMetAsp---LysGlyGlyIleIleArgGlnGluArgArg	334
Db	1588	-----GGCCCTGGCAAGATGATTCACAGGGGTGAGCACAGAGATCCGAGG	1635
QY	335	AspArgProTyr	338
Db	1636	GAGAGCGCGTAT	1647

```

RESULT 11
US-10-439-703-58
; Sequence 58, Application US/10439703
; Publication No. US20040018527A1
; GENERAL INFORMATION:
; APPLICANT: Chang, Jenny
; APPLICANT: O'Connell, Peter
; TITLE OF INVENTION: Differential Patterns of Gene Expression that Predict for Docetaxel
; TITLE OF INVENTION: Chemosensitivity and Chemoresistance
; FILE REFERENCE: HO-P2482US1/10205813
; CURRENT APPLICATION NUMBER: US/10/439,703
; CURRENT FILING DATE: 2003-05-15
; PRIOR APPLICATION NUMBER: US 60/381,141
; PRIOR FILING DATE: 2002-05-17
; NUMBER OF SEQ ID NOS: 91
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 58
; LENGTH: 1824
; TYPE: DNA
; ORGANISM: Human
US-10-439-703-58

```


This Page Blank (uspio)